

VOLUME 29

SEPTEMBER, 1951

NUMBER 5

Canadian Journal of Physics

***Editor:* G. M. VOLKOFF**

***Published by* THE NATIONAL RESEARCH COUNCIL
OTTAWA CANADA**

CANADIAN JOURNAL OF PHYSICS

This was formerly *Section A, Canadian Journal of Research*. The change to the new name took place January 1, 1951. The CANADIAN JOURNAL OF PHYSICS is published six times annually.

The CANADIAN JOURNAL OF PHYSICS is published by the National Research Council of Canada under the authority of the Chairman of the Committee of the Privy Council on Scientific and Industrial Research. Matters of general policy are the responsibility of a joint Editorial Board consisting of members of the National Research Council of Canada and the Royal Society of Canada.

The National Research Council of Canada publishes also the following Journals: *Canadian Journal of Botany*, *Canadian Journal of Chemistry*, *Canadian Journal of Medical Sciences*, *Canadian Journal of Technology*, *Canadian Journal of Zoology*.

EDITORIAL BOARD

<i>Representing</i>	
NATIONAL RESEARCH COUNCIL	ROYAL SOCIETY OF CANADA
DR. J. H. L. JOHNSTONE (<i>Chairman</i>), Professor of Physics, Dalhousie University, Halifax, N.S.	DR. G. M. VOLKOFF, Professor of Physics, University of British Columbia, Vancouver, B.C.
DR. OTTO MAASS, Macdonald Professor of Physical Chemistry, McGill University, Montreal, P.Q.	DR. J. W. T. SPINKS, Dean, College of Graduate Studies, University of Saskatchewan, Saskatoon, Sask.
DR. CHARLES W. ARGUE, Dean of Science, University of New Brunswick, Fredericton, N.B.	DR. H. S. JACKSON, Head, Department of Botany, University of Toronto, Toronto, Ont.
DR. A. G. MCCALLA, Department of Plant Science, University of Alberta, Edmonton, Alta.	DR. E. HORNE CRAIGIE, Department of Zoology, University of Toronto, Toronto, Ont.
<i>Ex officio</i>	
DR. LÉO MARION, Editor-in-Chief, Division of Chemistry, National Research Laboratories, Ottawa.	<i>Representing</i> THE CHEMICAL INSTITUTE OF CANADA DR. H. G. THODE, Department of Chemistry, McMaster University, Hamilton, Ont.
DR. H. H. SAUNDERSON, Director, Division of Information Services, National Research Council, Ottawa.	

Subscription rate: \$2.00 a year. All enquiries concerning subscriptions should be addressed to the CANADIAN JOURNAL OF PHYSICS, National Research Council, Ottawa, Canada. Special rates can be obtained for subscriptions to more than one of the Journals published by the National Research Council.



Canadian Journal of Physics

Issued by THE NATIONAL RESEARCH COUNCIL OF CANADA

VOLUME 29

SEPTEMBER, 1951

NUMBER 5

SPECTROSCOPIC STUDY OF SMALL SAMPLES IN A HOLLOW CATHODE DISCHARGE¹

By K. B. NEWBOUND AND F. H. FISH

Abstract

Operating characteristics of the hollow cathode discharge have been investigated qualitatively. A sample of less than $10\text{ }\mu\text{gm.}$ of an element may be used to study the hyperfine structure of the principal lines in an exposure of a few hours. The useful lifetime of the sample in the discharge is about 10 hr., and up to 80% of the sample can be recovered after use.

Introduction

The cooled hollow cathode type discharge has proved a very useful light source for the study of the hyperfine structure in line spectra (4). In general, a few milligrams of sample material are required. However, Schüler and Schmidt (3) found that, with a microgram of thallium in a liquid air cooled hollow cathode, the hyperfine structure of the green line (5350Å) could be photographed. McNally, Harrison, and Rowe (1) were able to detect 10^{-8} gm. of fluorine in an analysis problem. In view of the interest in the nuclear spins and "g values" of radioactive isotopes, it was felt that further investigation of the practicability of using small amounts of sample in the hollow cathode was desirable.

Apparatus

The experiments to be described made use of (i) a hollow cathode discharge tube in conjunction with (ii) a gas circulating and purification system, (iii) a power supply to energize the discharge, and (iv) a quartz spectrograph for the examination of the light emitted by the discharge.

The spectrograph used was a Hilger E-1 equipped with quartz optics. A quartz lens with a focal length of 5 cm. was used to focus the discharge on the slit of the spectrograph. Eastman IIIF or IIIG emulsions were used to record the spectra.

The power supply and gas circulating system were of conventional design.

Hollow Cathode Design

The discharge tube (cf. Fig. 1) consisted of water cooled anode and cathode sections separated by a Pyrex glass spacer sealed in place with red wax. Brass

¹ Manuscript received February 2, 1951.

Contribution from the Department of Physics, University of Alberta, Edmonton, Alta. This research was supported by a grant from the National Research Council of Canada, and one of the authors (F.H.F.) was holder of a Bursary under the National Research Council of Canada (1949-50).

was the principal material used in the construction, while the actual electrodes were aluminum inserts, which could be accurately positioned in the anode and cathode sections. In order to simplify the diagram, soldered joints have not been shown.

Noteworthy features in the discharge tube are:

(1). The use of aluminum electrodes. It is desirable that the electrodes be coated with some refractory material to decrease the spectrum of the electrode material. In the case of iron this is accomplished by artificially oxidizing the surfaces in contact with the discharge. Aluminum, of course, requires no special treatment, since the surfaces are oxidized immediately on contact with air.

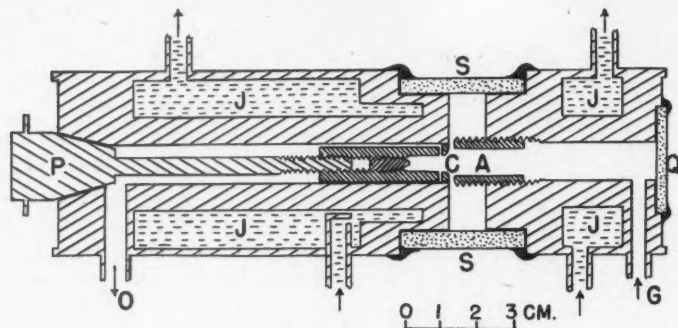


FIG. 1. Hollow cathode mount. A, anode; C, cathode; G, gas inlet; J, J, water jackets; O, gas outlet; P, tapered plug for withdrawal of cathode; Q, quartz window; S, pyrex spacer.

(2). The use of a tapered plug coated with vacuum grease provides for the rapid and convenient replacing of loaded cathodes in the cathode section. This design was possible since water was used for cooling rather than liquid air. The use of the latter produces difficulties when joints are sealed using high vacuum greases and waxes, since these crystallize at low temperature leading to leaks in the apparatus.

(3). Placing the inlet tube for the circulating gas near the quartz window produces a stream of gas through the anode toward the cathode, which tends to prevent sputtered cathode material from reaching the window and rendering it opaque.

(4). The aluminum cathode finally adopted consisted of an aluminum tube of 5 mm. inside diameter, threaded to receive an aluminum insert, and provided with six radial holes 0.8 mm. in diameter set around the circumference about 1 mm. from the front face. These holes were connected by a groove which joined an opening leading to the rear of the cathode. The circulating gas entered the front face of the cathode and left by means of these holes. This design produces (2) a gas "window" near the front face of the cathode, which tends to prevent particles from the cathode load from reaching the anode,

being deposited there, and so being lost to the discharge. The cathode insert, which formed the back surface of the hollow cathode, was in the shape of a truncated cone, the flat tip being 2 mm. in diameter. It was on this tip that the load material was placed. Earlier models of the cathode were constructed in one piece, the discharge region being simply a hole drilled in an aluminum rod to the desired depth. The characteristics of the two types are discussed below.

Experimental

Loading of Samples

When small amounts of sample material are to be used in a hollow cathode, it becomes important to examine the effects of different locations of the sample. Using the one piece cathode of the type used by Schüler and Gollnow (2), it was found that the sample was most effectively excited when placed on the wall near the front. Loading was carried out using a small hypodermic needle to deposit known small amounts of solution. Difficulty was experienced in placing the solution in the desired location and then maintaining it there during evaporation. It was convenient, therefore, to use the cathode insert described above. When the sample was placed on the tip and evaporated slowly the use of the insert provided similar discharge characteristics to those of cathodes loaded near the front. The gas "window" formed by the circulating gas in front of the cathode opening did not appear to be appreciably affected when the distance between the tip of the insert and the front of the cathode was about 1.0 cm. When this distance was 5 mm., the lines of the load material were intense initially, but disappeared almost entirely after about 15 min. operation, indicating that the effectiveness of the gas "window" had been greatly reduced. A distance of 1.5 cm. was found to produce a discharge almost completely lacking in lines of the load material. Subsequent discussion will therefore deal only with a tip-to-front separation of 1.0 cm.

Operating Conditions

In order to avoid the anomalous fall in cathode potential, estimated by Tolansky (4) to take place when the current density exceeds 30 ma. per cm.², the discharge current used in this research was 15 ma. Smaller currents were found to produce discharges whose light emission was lacking in the load lines.

The intensity of the light emitted by the hollow cathode was found to increase markedly as the anode-cathode spacing was reduced. A spacing of less than 1 mm. was therefore used for the bulk of the work. The pressure of the circulating gas was adjusted to give maximum light intensity in all trials. Variation in the rate of gas circulation within the limits available with the diffusion pump did not appreciably affect the discharge.

Discharge Characteristics

The assessment of discharge characteristics in this work was limited to the intensity of the load lines with respect to lines of the circulating gas, and to

the length of time over which the load lines could be observed. Materials used to load the cathode were sodium chloride, magnesium chloride, copper sulphate, zinc chloride, and silver nitrate. In each case the load applied was 10^{-5} gm. of the metal. It was found that the more intense lines of all these elements had an intensity comparable to the helium line at 2829\AA after five hours of operation, providing that the discharge had operated continuously. Magnesium and copper lines were still detectable in a one hour exposure after nine hours of previous operation. On observing the load lines after resumption of the discharge, it was found that their intensity was greatly reduced. Apparently a large fraction of the sample condenses on the less active portions of the system when the discharge is stopped and is not released when it is restarted.

Since equal masses of the metals were used in all cases, the numbers of atoms available for excitation were greater for the lighter elements than for the heavier ones. For instance there were about 2.5 times as many magnesium atoms available as copper atoms. Yet the copper spectrum was in general the more intense. Examination of the insert tips after use indicated that the copper remaining had been reduced to metallic form, appearing as a thin film on the tip. The sodium and magnesium tips were somewhat whitish indicating the presence of a salt of some kind. From the absence of any observable film like that of a salt or oxide, the silver and zinc tips probably also had metallic films after use.

It appears, therefore, that the intensity effects are related to the electropositeness of the load metal with respect to aluminum. The latter is electropositive with respect to copper, silver, and zinc. Hence its presence tends to reduce salts of these to metallic form. On the other hand, aluminum is electronegative with respect to sodium and magnesium, and these elements remain as salts or perhaps oxides. Thus to enhance the intensity of the sample lines, conditions should be arranged so that the sample element is not chemically combined.

Recovery of Sample

Using a load of radioactive material, it was found that after three hours' operation with a discharge current of 15 ma. about 30% of the sample was lost from the insert. Anode and cathode each showed about 10% of the initial count, while nothing above background could be located in the rest of the circulating system. Thus, it should be possible to recover up to 80% of the load material from the anode, cathode, and insert, even after 8 or 10 hr. of operation.

Suitability for Interferometry

The principal lines of the elements with which the cathode was loaded gave satisfactory blackenings on the photographic plates in a 1 hr. exposure when no interferometer was used. Experience with the Fabry-Perot interferometer (mirrors aluminized to 80% reflectivity) indicates that, when interferometric studies are to be made, this exposure time should be increased to between 5

and 10 hr. This time is of the order of the useful life of the sample in the light source, and could be reduced by the use of more sensitive photographic plates than those used in these experiments.

References

1. McNALLY, J. R., HARRISON, G. R., and ROWE, E. J. *Optical Soc. Am.* 37: 93. 1947.
2. SCHÜLER, H. and GOLLNOW, H. *Z. Physik*, 93: 611. 1935.
3. SCHÜLER, H. and SCHMIDT, T. *Z. Physik*, 96: 485. 1935.
4. TOLANSKY, S. *High resolution spectroscopy*. Methuen & Co., London. 1947.

THE EFFECT OF ION BOMBARDMENT ON THE EMISSION FROM OXIDE COATED CATHODES¹

BY P. A. REDHEAD

Abstract

Some preliminary measurements of the decay of emission caused by ion bombardment of oxide coated cathodes are presented. It is shown that cathode life increases with emission density if ion bombardment is prevented.

1. Introduction

The constant evaporation of emitting material from the surface of an oxide cathode results ultimately in failure of the cathode emission, thus ending the useful life of the tube. Frequently cathodes lose their emission long before all the active material of the cathode coating becomes exhausted; this second phenomenon is termed cathode "poisoning" and is caused by foreign atoms or ions entering the cathode coating and altering its emitting properties. An experimental study of the changes in emission caused by bombardment of the cathode surface by positive ions formed in normal tube operation has been made, and is the subject of this paper.

Ions may be created within vacuum tubes by three main processes:

- (a) Ionization by collision of the electron stream with residual gases in the tube.
- (b) Electron bombardment of contaminated electrodes ("anode effect"). If the energy of the impinging electrons is in excess of the heat of formation of the contaminating material, this material will decompose to form ions (3).
- (c) When the cathode is operated under temperature limited conditions, oxygen is evolved as neutral atoms or molecules (1). This oxygen may be ionized by electron collision and the positive oxygen ions will return to the cathode.

Processes (a) and (b) are the most important.

It is known that oxygen ions are the chief cause of ion poisoning, pressures of 10^{-4} mm. of oxygen being sufficient to decrease the emission by several powers of ten (6). Oxygen ions entering the cathode coating are presumed to combine with the free barium in the coating and thus reduce the density of the barium impurity centers. The saturated emission density from the cathode is given by the familiar equation

$$j_s^0 = B(1 - r)T^{5/4}n_b^{1/2} \exp \left[\frac{-e}{kT} \left(\chi + \frac{\Delta e}{2} \right) \right] \quad (1)$$

¹ Manuscript received February 28, 1951.

Contribution from the Radio and Electrical Engineering Division, National Research Council, Ottawa. Issued as N.R.C. No. 2484.

where B is a constant, r the reflection coefficient, T the temperature, and n_b the density of impurity centers. The surface work function χ is a constant and the energy interval $\Delta\epsilon$ is only slightly dependent on n_b . Equation (1) shows how the emission is affected by ion bombardment on the assumption that it causes a reduction in the impurity center concentration, n_b . Other factors besides the change in impurity center density are operative in causing this form of emission decay.

Oxide cathodes are most sensitive to poisoning when they are operated under temperature limited conditions. This is probably due to the evolution of oxygen from the surface of the cathode, mentioned in (c) above. It is fortunate, since tubes are normally operated space-charge limited, that under space-charge limited conditions poisoning occurs slowly. However under pulsed conditions tubes are frequently operated temperature limited and the effect of ion bombardment is clearly apparent when using long pulses.

When emission can be recovered from a poisoned cathode by operating for a short period at an increased cathode temperature, or by removing the source of poisoning material, the poisoning is called reversible. In this case the cathode recovers by ejecting the poisoning atoms or ions and the emission finally returns to the value it had before poisoning started.

Emission failure of tubes in electronic equipment is due either to loss of active material, as mentioned in the first paragraph, or to a form of irreversible poisoning which is closely associated with reversible poisoning. Reversible poisoning occurs continuously during the life of oxide cathodes, the cathode being poisoned and reactivated simultaneously. It is postulated that there is a progressive failure of minute areas of the cathode surface to reactivate, thus the effective emitting area of the cathode decreases with time. This causes the commonly observed phenomenon of decreasing anode current and mutual conductance.

By operating tubes at potentials less than the lowest ionization potential of any residual gases in the envelope, and also below the breakdown potentials of any material deposited on the electrodes, it is possible to ensure that no ions are formed. The effect of positive-ion bombardment on emission has been studied by operating two tubes in the same envelope (to ensure equal residual gas pressures), one tube being operated with an anode voltage below the lowest ionization potential, and the other at a higher potential. Thus the first tube shows the decay of emission due to causes other than ion bombardment and the second diode shows the decay due to all causes.

2. Long Time Effects

2.1. Temperature Limited Conditions

The decay of emission under these conditions is quite rapid, a decrease to 50% of the initial emission may occur within a few hours (5).

The decay of temperature limited emission was first observed using com-

mercial tubes. Double triodes with separate cathodes were employed to ensure that the pressure of residual gases would be identical in the two tubes tested under different conditions. Batches of 10 tubes were taken, one half of the double triode was diode connected by connecting the grid and anode, while the other half was triode connected. The diode section was operated at 4 v. on the anode, the triode section was operated at $V_a = 300$ v. The triode grid was held at space potential. The temperature of the cathodes was such that both sections were temperature limited, and thus the initial anode current through both diode and triode was substantially the same. The only difference in operating conditions in the two cases was that the cathode of the triode section was subject to ion bombardment, while the diode cathode was not.

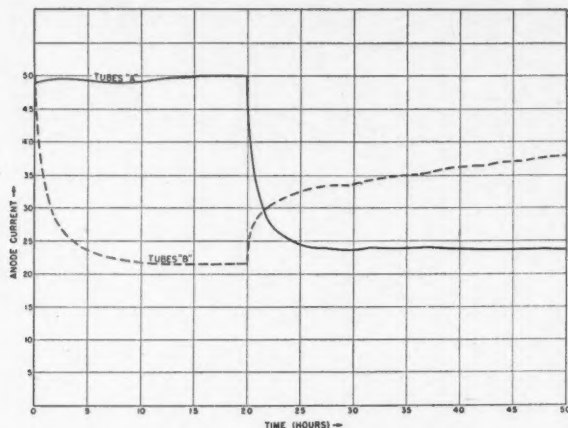


FIG. 1. Variation of temperature limited anode current (in arbitrary units) with time. Initially Tubes A were diode connected and Tubes B were triode connected.

The results of operating a batch of 10 tubes under these conditions are shown in Fig. 1, which is representative of all the runs under these conditions. From time zero to 20 hr. Tubes A were operated as diodes and Tubes B as triodes. Over this portion of the curve the diode emission remained sensibly constant while the triode emission (i.e., the emission from the cathodes subject to positive-ion bombardment) decreased rapidly at first and finally approached a steady state. At time 20 hr. the connections to the two batches of tubes were reversed and Tubes A were triode connected, while Tubes B were diode connected. The decay of Tubes A is similar to the earlier decay of Tubes B. Tubes B show a slow reactivation which continued for about 50 hr. before reaching a steady state; this steady state value was only 75% of the initial value of emission. The initial emission could be regained by operating the tubes at an increased cathode temperature for a few minutes without drawing current, showing that the cathode poisoning for this relatively short operating time was entirely reversible.

The conditions for the test shown in Fig. 1 were

$$\begin{aligned} V_a(\text{diode}) &= 4 \text{ v.}, \\ V_a(\text{triode}) &= 290 \text{ v.}, \\ \text{Cathode temperature} &= 800^\circ\text{K}^*. \end{aligned}$$

Numerous types of cathode core materials and coating material have been tested in this manner. They were all found to give very similar results.

2.2. Space-charge Limited Conditions

Oxide cathodes operated space-charge limited are considerably less sensitive to poisoning by ion bombardment than when operated temperature limited. This may be due in part to the fact that no oxygen is evolved by a space-charge limited cathode.

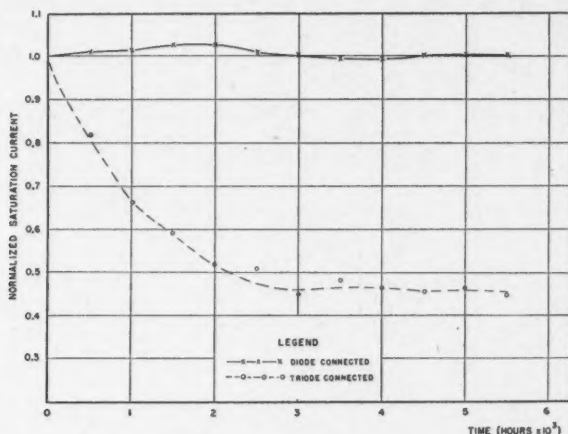


FIG. 2. Decay of saturated emission under space-charge limited operating conditions.

A batch of 10 double triodes was connected as in the previous experiments. The diode connected section was operated at 1.2 v. on the anode and the triode at 200 v.; the grid voltage of the triodes was adjusted to make the triode anode current initially equal to the diode current. The cathodes were operated at approximately 1050°K . Saturated emission was measured at intervals by lowering the cathode temperature to about 750°K . and measuring the anode current. The emission from the diode section remained fairly constant for at least 6000 hr., while the emission from the triode section fell until it reached a steady value at about 4000 hr. Fig. 2 is typical of results obtained on several tests. The normalized saturation current refers to the current at zero time as unity.

The presence of a high percentage of silicon in the cathode core material causes a rapid decay of emission owing to the formation of a high resistance

* All cathode temperatures quoted were measured with an electronic temperature bridge operating in the retarding field region, having an accuracy better than $\pm 2\%$ (2, 4).

interface between the cathode core and the coating material. Fig. 3 illustrates such a case. In this example the core contained 0.25% silicon and the tubes were operated at a low current density (10 ma. per cm.²), both of these factors being conducive to the growth of a high resistance interface. Both triode and diode connected sections are equally affected by the growth of the interface.

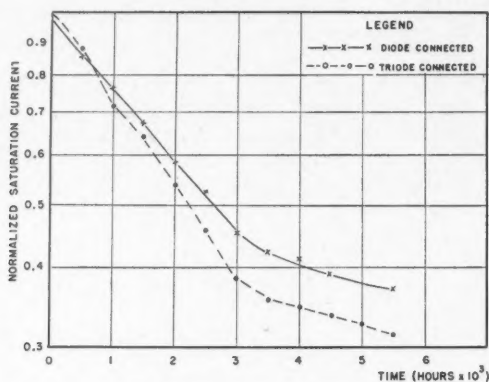


FIG. 3. Decay of saturated emission under space-charge limited operating conditions when an interface is formed.

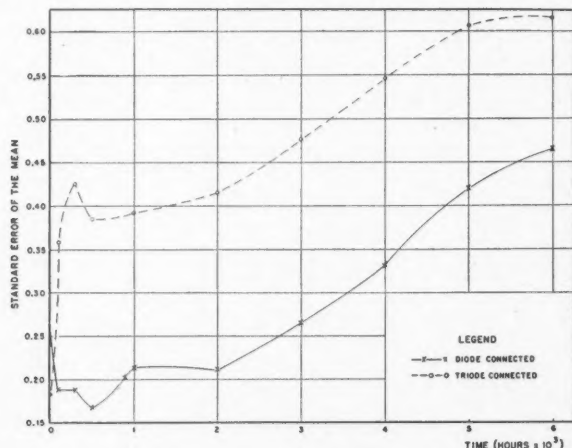


FIG. 4. Variation with time of the standard error of the mean of the anode current of a batch of 10 tubes.

Fig. 4 shows the variation of space-charge limited anode current amongst a group of 10 tubes expressed as the standard error of the mean. Initially there was a greater dispersion amongst the diode connected sections than amongst the triode sections. Within the first 500 hr. the dispersion amongst the diode

connected batch decreased rapidly while the triode dispersion increased. After 1000 hr. both dispersions increased at the same rate.

3. Effect of Emission Density

It is usually assumed that increasing the emission density from an oxide cathode will result in reduced cathode life. However, experiments have shown that cathode life increases with increasing emission density provided that the anode voltage is below ionization potential (7).

Special experimental tubes were used in these tests containing two diodes in one envelope. The cathode sleeve of Grade "A" nickel (0.05% silicon) was

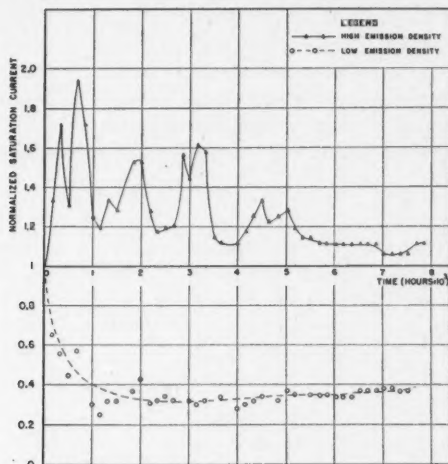


FIG. 5. Variation with time of the saturated emission from two diodes operated at different space-charge limited current densities.

coated to a thickness of 1/20 mm. Two anodes of different diameter (1.9 mm. and 4.5 mm. respectively) were used so that different space-charge limited currents could be drawn to the two anodes which were maintained at the same potential. The temperatures of the cathodes of the two diodes were measured under operating conditions with an optical pyrometer and were found equal to within the accuracy of measurement.

Fig. 5 shows the normalized saturation current as a function of time for one batch of tubes. A measure of the saturated emission was obtained by depressing the cathode temperature. The data of Fig. 5 were obtained from a batch of 10 tubes with an anode voltage of 9 v. and a cathode temperature of 950°K. The initial space-charge limited current densities drawn to the large and small anodes were 50 ma. per cm.² and 625 ma. per cm.² respectively. It will be seen that the saturated emission of the diode operated at low emission density de-

creased rapidly to a steady value of about 35% of the initial emission. The saturated emission from the diode with a high emission density varied considerably in the first 5000 hr. and finally remained constant at about 110% of its initial value. These general results were confirmed by a number of batches of tubes operated in the same manner.

It has been found that the difference in rates of change of emission is greater at low cathode temperatures. As the cathode temperature was increased the difference in the two rates of decay decreased until at 1000°K. the difference is small. At these high cathode temperatures the rate of decay of emission of the low emission density cathode was always a little greater than the rate of decay for the high emission density cathode.

It is postulated that increasing the emission density causes an increase in the electrolytic production of free barium within the coating and thus enhances the emission. With voltages above ionization potential the increased current density creates a large ion current which poisons the cathode and obscures any effect due to increased electrolytic action. However at low anode voltages no ion poisoning occurs, and cathode life is found to increase with emission density.

At low cathode temperatures the production of free barium by thermal processes is reduced and thus the electrolytic production of free barium becomes the predominant process. This is evidenced by the large difference in rates of emission decay between low and high emission density cathodes at low cathode temperatures. As the cathode temperature is increased, and the electrolytic process becomes less important in relation to the thermal process, the difference between the rates of decay diminishes.

4. Conclusions

It has been shown that positive-ion bombardment is one of the chief causes of the decay of emission from oxide coated cathodes. No satisfactory fit to the observed decay curves can be obtained on the basis of the Wilson semiconductor theory assuming that the poisoning process is due only to a reduction in the impurity center concentration. Thus it appears that this form of emission decay is due to two or more simultaneous processes. It is suggested that the decay of emission is caused by a reduction of the number of impurity centers in the cathode coating and by sputtering of the cathode surface by the relatively heavy ions. It is suggested that the final steady value of emission represents an equilibrium between the poisoning and reactivation processes.

It has also been shown that the life of an oxide cathode is not diminished when the emission density is increased by a factor of as much as 20, provided that no ionization occurs within the tube.

Investigations are continuing to clarify the mechanism of ion bombardment poisoning and will be published later. It was felt that a brief description of the preliminary experimental results would be useful at this time.

References

1. BECKER, J. A. Phys. Rev. 34: 1323. 1929.
2. GAUDIN, P. and CHAMPEIX, R. Compt. rend. 229: 545. 1949.
3. JACOBS, H. J. Applied Phys. 17: 596. 1946.
4. McNARRY, L. R. To be published.
5. METSON, G. H. and HOLMES, M. F. Nature, 163: 61. 1949.
6. REIMANN, A. L. and MURGOCI, R. Phil. Mag. 9: 440. 1930.
7. WAGENER, S. Nature, 164: 357. 1949.

THE MACROMOSAIC STRUCTURE OF TIN SINGLE CRYSTALS¹

BY E. TEGHTSOONIAN² AND BRUCE CHALMERS³

Abstract

Single crystals of high purity tin grown from the melt by a modified Bridgman method are shown to be partitioned into bands, or striations. A difference of orientation exists between the striations such that the lattice of one striation may be brought into coincidence with the lattice of its neighbor by a pure rotation about an axis parallel to the specimen axis. The properties of the striations have been shown to be dependent on both the rate of growth and the crystallographic orientation relative to the direction of heat flow. A tentative explanation of the origin of this effect is advanced in terms of the formation of edge type dislocations from the condensation of vacant lattice sites.

Introduction

Mosaic structures occurring in metal or other crystals may be divided broadly into two main categories: micro- or macromosaics. Micromosaics are those which occur on an order of magnitude of about one micron, while macromosaics are usually of a visually observable order of magnitude.

Considerable attention has been paid in the past to micromosaic structure. This has its origin in the theoretical work of Darwin (7) who suggested the existence within a crystal of a mosaic block structure of slightly differing orientation from block to block in order that observed and calculated X-ray reflection intensities should be in closer agreement. Since Darwin's work much experimental and speculative work has been done which is summarized in a paper by M. J. Buerger (1), who has also pointed out the existence of what he called "lineage structure". This was a term applied to a crystal that is partitioned into branches or "lineages", each lineage being descendent from a common nucleus, and having a slightly different orientation from its neighboring lineage.

This paper describes some work that has been done on a type of lattice imperfection that produces a mosaic structure of the macroscopic type. Because of their macroscopic character, these mosaics are readily observable and their behavior and properties may be examined by relatively simple techniques.

A careful study of these mosaics has been undertaken in single crystals of high purity tin. They do not, however, occur solely in tin. They have been observed in lead, zinc, copper, and silver single crystals as well. Because of its low melting point and relative ease of handling, tin was chosen for a more detailed study.

¹ Manuscript received February 1, 1951.

² Contribution from Department of Metallurgical Engineering, University of Toronto, Toronto, Ont.

³ Graduate Student, Department of Metallurgical Engineering, University of Toronto.

⁴ Professor of Physical Metallurgy, Department of Metallurgical Engineering, University of Toronto.

In 1935, Greninger (9) reported the results of work that he had done on the lineage structure of copper single crystals. He grew single crystals of copper by the Bridgman method and examined the fine structure obtained in the X-ray back-reflection Laue spots. From this fine structure he was able to show that the single crystal was partitioned into lineages between which there existed a difference of orientation of from five minutes to two degrees. He also showed that these lineages had a predominantly common $[110]$ direction and that they were separated by a predominantly common $\{111\}$ plane.

The present work on tin somewhat resembles that of Greninger, but differs from it in three rather important respects. First, the method of growing the single crystal specimens is such that crystals of any desired orientation may be grown. Even though Greninger's method yielded a high percentage of good single crystals, it did not permit control of orientation. Second, the method of growth yields single crystal specimens in which the lineage structure is visually observable. This provides a very powerful additional way of examining the behavior of the structure under various conditions of growth. Third, the rate of growth may be controlled and determined at leisure after the crystal is completely grown.

Method and Materials

Thus far, only high purity tin has been used for the experiments. This tin is labelled "Chempur", supplied by Messrs. Johnson, Matthey of London, England. Manufacturer's spectroscopic analysis shows this to be 99.987% pure with the principal detectable impurities being Pb, Bi, Fe, Sb, and Cu.

The method used for growing single crystals of tin was developed by Chalmers (4). The tin is melted in a graphite container, called a boat, which is generally of rectangular cross section. Typical dimensions are 20 cm. long, 2 cm. wide, and 1/2 cm. deep. This boat is clamped in a horizontal position, extending into a small electric furnace made of a few turns of Chromel A wire. The furnace is mounted on a carriage which may be driven at any desired speed along tracks parallel to the axis of the boat by means of a geared drive.

A seed crystal of known orientation is placed in the boat at the end remote from the furnace and a junction is made between this seed and the molten tin. The crystal is then made to grow by moving the furnace along the tracks in a direction away from the seed. This will cause the interface separating solid from liquid to advance into the liquid and eventually all the liquid tin will have solidified. If all the operations are done carefully, the liquid will solidify into a single crystal of the same orientation, relative to the axis of the specimen, as that of the seed.

Obtaining a seed crystal to start with presents no great difficulties. To do this, some tin is melted into a boat as previously described. Then, using a glass rod, the liquid tin is drawn out from one end to as fine and narrow a point as possible. By slowly moving the furnace in a direction away from the pointed end, the molten tin is made to solidify. Conditions such as this are very favor-

able for the nucleation, at the fine point, of a single crystal which will grow to the exclusion of other crystals.

Having obtained a single seed crystal in this manner, crystals of any desired orientation may be grown by orienting the seed in an appropriate direction when making the junction with the melt. Single crystals of tin have been successfully grown by this method at rates as high as 3 or 4 cm. per min. At rates greater than 4 cm. per min. there is a strong tendency for stray crystals to form.

A valuable feature of this technique is that periodically the liquid end of the system may be agitated. This produces a ripple which freezes at the instantaneous position of the solid-liquid interface. From these ripples the shape of the interface and the rate of growth of the crystal may be determined at leisure.

In tin, the structure, which has been called "striations" becomes apparent after a short etch in a solution of 10% ferric chloride, 10% hydrochloric acid, and 80% water.

Experiments

It has been pointed out that the technique used should allow growth of crystals of any desired orientation. In practice, however, the success with which single crystals may be grown is, to a considerable extent, dependent upon the orientation. It has been found that the "easy" direction of growth is that in which the direction of longitudinal heat flow is parallel to the [110] direction. Accordingly, most of the preliminary experiments were done on crystals in which the [001] direction was perpendicular to the top surface of the crystal and the [110] direction was parallel to the specimen axis.

When crystals of this orientation are etched, they show on the top free surface a series of fine bands running down the length of the specimen parallel to the specimen axis; see Fig. 1. These bands are quite strikingly visible when the surface is examined with light at an appropriate angle, owing to the difference in reflectivity in each band. The widths of the bands, or striations, vary considerably, but are generally in the range from 1/4 mm. up to 2 mm.

In all but a few cases, striations are not visible on the bottom surface of the crystal. A possible explanation for this is given below.



FIG. 4. Schematic diagram of crystal arrangement in boat.

Usually crystals are grown from seeds that are narrower than the crystals themselves, a smooth junction being made by means of a graphite insert as shown in Fig. 4.

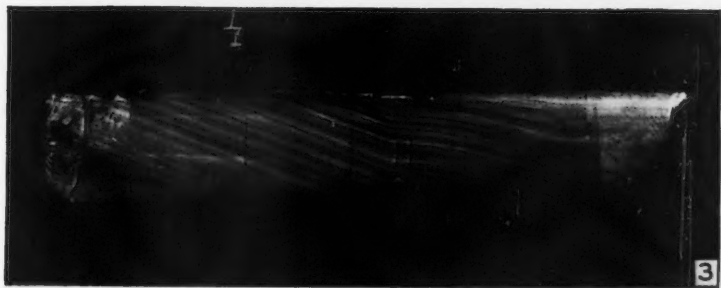
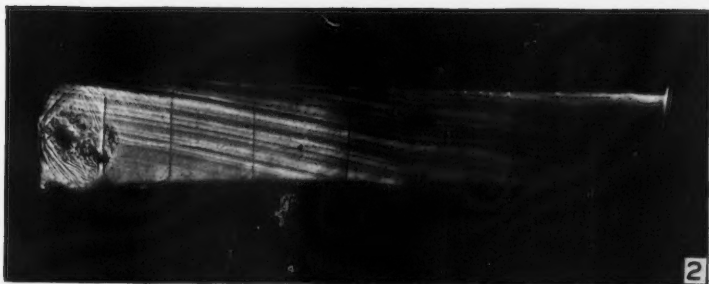
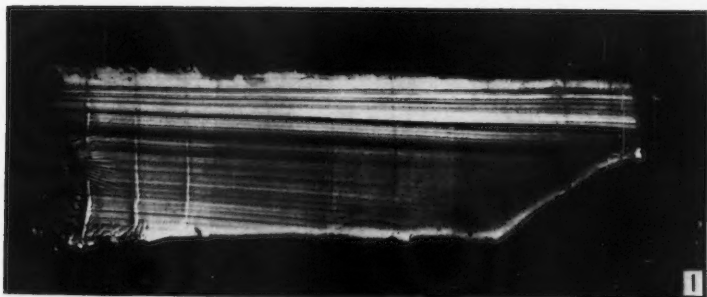
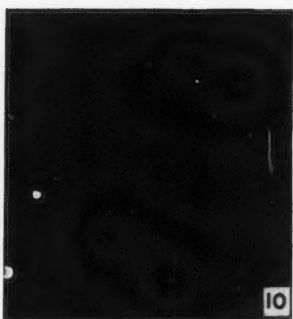
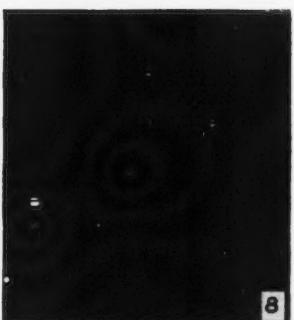
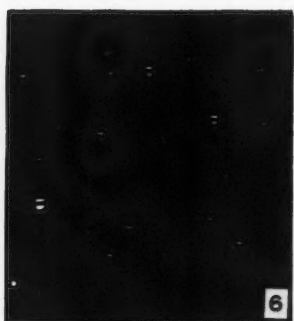
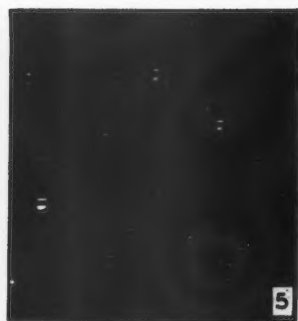


FIG. 1. Top surface of single crystal showing striations. *C* axis perpendicular to plane of paper, *A* and *B* axes at 45° to specimen axis.

FIG. 2. As Fig. 1 with 15° rotation about *C* axis. Lines crossing the specimen are at one minute intervals.

FIG. 3. As Fig. 1 with 26° rotation about *C* axis.



FIGS. 5 TO 10. X-ray pictures along one pair of new striations taken at 2 mm. intervals. Note gradual approach of the spots in each multiple.

Under these conditions, it has been found that the striations present in the seed continue through into the crystal. This is not possible for that portion of the crystal that is in line with the graphite insert. In this region, new striations form independently from the seed. As can be seen in Fig. 1, these new striations generally start at some distance from the sloping edge at the graphite insert.

Crystal Structure in Striations

An ordinary back-reflection X-ray Laue picture from a single crystal consists of single spots arranged on hyperbolas. All these spots lying in one hyperbola are due to reflections from crystallographic planes in one zone.

In the back-reflection pictures taken from the surface of a crystal showing striations, the pattern shows each reflection to be multiple instead of single. The multiplicity of each reflection depends on the width of the striations and the size of the X-ray beam at the surface of the crystal. Thus, the multiplicity is found to be equal to the number of striations covered by the X-ray beam. Fig. 5 shows the nature of this multiplicity.

All the spots in a given multiple are due to reflections from planes of the same (hkl) but each tilted at a slightly different angle to the direction of the incident X-ray beam. From the regular manner in which the multiplicity develops for all reflections, it is apparent that there is an orientation difference between the lattices of neighboring striations consisting of a pure rotation about a common direction. This rotation, which would bring the lattice of one striation into coincidence with that of its neighbor, may be measured using Greninger's (10) chart, and has been found to be in the range from approximately $1/2^\circ$ up to 4° .

Crystals which are grown with the [110] direction parallel to the specimen axis and the [001] direction perpendicular to the top surface always show striations parallel to the specimen axis. In all these specimens, back-reflection X-ray pictures show that the difference of orientation between neighboring striations is a rotation about the [110] direction as axis.

Series of X-ray pictures have been taken across the width of a crystal over a distance of 15 mm. in which there might be anywhere up to 30 or 40 striations. These have shown that the orientation at any point in this region always lies somewhere in a range bounded by 5° or 6° . This indicates that the rotation of the lattice in going from one striation to the next is not always in the same direction. Rather, there are about as many rotations in one direction as there are in the other.

This has been shown more precisely by taking a series of X-ray pictures at intervals across the width of the specimen equal to the diameter of the X-ray spot. Then, by superimposing neighboring pictures, a set of continuous measurements of rotations of the lattice from striation to striation may be made, both in magnitude and direction. These results show that the sum of all the

rotations in one direction is very nearly equal to the sum of all the rotations in the other direction.

Striation Distribution

Two experiments have given conclusive evidence that the striations extend throughout the whole volume of the crystal.

First, a crystal of approximately 5 mm. thickness was started growing in the usual manner. With the solid-liquid interface advancing slowly into the liquid, the boat was suddenly tipped so that the portion still liquid ran away. This decanting process leaves exposed what was essentially the solid-liquid interface at the time of decanting. Several back-reflection X-ray pictures taken at random positions of this decanted interface all showed multiple spot patterns, even though no visible striation structure was present. The direction of the multiple spots again showed a difference of orientation consisting of rotation about the [110] direction.

Second, a crystal which showed a very marked striation pattern on its upper surface and virtually no striations on its lower surface was subjected to a very strong electrolytic etch. The electrolyte used was a 10% solution of hydrochloric acid in water and the current density was about 0.1 amp. per sq. cm. The over-all effect was to reduce the crystal to about one-half of its original thickness. At all stages of etching, visual examination showed a striation pattern on each newly exposed surface. Back-reflection X-ray pictures from the final surface showed multiple spot patterns similar to those from the original un-etched surface.

Thus, there can be no doubt that the imperfections that give rise to striations extend throughout the whole volume of the crystal.

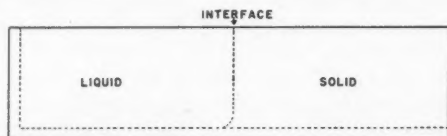


FIG. 11. Schematic diagram of shape of interface.

It is interesting to note that, after etching, the newly exposed surfaces on the bottom side of the crystal also showed striation where none had been seen before. This would suggest that the solid-liquid interface is quite sharply curved just at the bottom of the boat, where contact is made with the graphite owing to the higher thermal conductivity of the graphite than of the tin. Previous observations had shown that, when the rate of growth becomes sufficiently high, the trace of the interface at the surface becomes curved concavely towards the liquid. When this was the case, the striations no longer formed in a direction parallel to the specimen axis but instead they converged showing a tendency to grow perpendicularly to the interface. Thus, if the inter-

face just at the bottom of the boat was curved, the striations would tend to develop perpendicularly to this interface and would grow up into the crystal rather than parallel to the surface. (See Fig. 11.) The etching, having removed the thin layer where this had happened, revealed the striations in layers further up, which had developed in the normal way.

Striation Width

For crystals of the previously described orientation, it had been observed qualitatively that the width of the striations increased as the rate of growth decreased. An attempt was made to correlate striation width and growth rate in a quantitative way. A number of factors tend to make this correlation rather difficult:

1. The difficulty of growing crystals under exactly controlled and identical conditions,
2. The fact that striations in any one crystal vary considerably in width,
3. The difficulty in simultaneously observing and measuring the striations.

The first of these may be reduced by maintaining thermal conditions as steady as possible and waiting for equilibrium to be established between the solid and liquid before starting the timed growth. The rate of growth is then very nearly the same as the rate of motion of the furnace. To overcome the second difficulty, the average of a large number of measurements is taken. Also, instead of listing average widths for single growth rates, the average width for a range of growth rates is given. To measure the striations, it was eventually found best to photograph the crystal and make measurements on an enlarged print.

TABLE I

Rate interval, mm. per min.	Average width, mm.
0-1.0	1.20
0.5-1.5	1.05
1.0-2.0	0.85
1.5-2.5	0.75
2.0-3.0	0.60
2.5-3.5	0.50
3.0-4.0	0.45
3.5-4.5	0.40
5.5-6.5	0.35
9.0-10.0	0.25

Measurements were carried out on crystals grown at rates of from 10 mm. per min. down to less than 1 mm. per min. Table I shows the results. This table shows a trend towards a constant average width as the rate of growth increases. This is not strictly true. Owing to the curvature of the solid-liquid interface at the higher rates of growth mentioned before, the striations rapidly converge and become so fine that they cease to possess individual identity and any width measurement becomes impossible.

Stability

The stability of the imperfections giving rise to the striations has been tested by subjecting a specimen to successive 120 hr. and 240 hr. anneals at 227°C. (melting point of tin is 232°C.). From visual examination and from X-ray pictures taken from the same positions of the surface before and after the anneal, it has been shown that there is no change in either boundary positions or orientation difference. This shows that the structure of the imperfection is extremely stable.

Orientation Effect

All the experiments described so far have been done with crystals oriented with the [110] direction parallel to the specimen axis, and the [001] direction perpendicular to the top surface. In order to determine the effect of the crystallographic orientation on the formation of striations, it was necessary to grow crystals with the direction of heat flow parallel to other crystallographic directions.

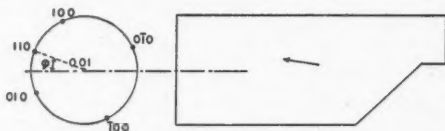


FIG. 12. Schematic diagram showing orientation of crystal by simplified stereographic projection. Arrow indicates direction of striations.

With the [001] direction still perpendicular to the top surface of the specimen, we can specify an angle φ between the [110] direction and the specimen axis (direction of longitudinal heat flow) (See Fig. 12). Thus, for the crystals previously described, $\varphi = 0$.

Crystals were grown in which the angle φ was progressively increased in steps of approximately 10 degrees from 0° to 45°. For each orientation, crystals were grown at various rates. In all cases, the general results are the same. At low rates of growth, the striations formed are virtually parallel to the specimen axis. As the rate of growth is increased, the striations form at progressively greater angles to the specimen axis. Figs. 2 and 3 show this effect. The maximum angle which the striations make with the specimen axis increases as the angle φ increases but not so rapidly. Thus, there is a tendency at high rates of growth for the striations to be parallel to the [110] direction. At these high growth rates, the angle between striations and the [110] direction increases as the angle between the [110] direction and the specimen axis increases. The results are shown graphically in Fig. 13.

Here, the results are plotted as the angle between the striations and the specimen axis. A number of features should be noted in these curves.

1. As the angle φ is increased, the rate of growth which is necessary to produce constant angle of formation of the striations is increased.

2. As the angle φ is increased, the maximum angle between striations and the specimen axis is also increased.

Back-reflection X-ray pictures indicate that the difference of orientation between neighboring striations is always a rotation about an axis parallel to the specimen axis, regardless of the angle between the striations and the specimen axis.

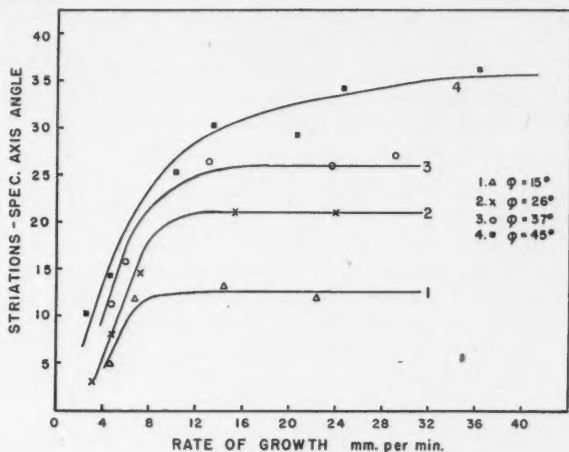


FIG. 13. Curves relating direction of striations and rate of growth for different orientations.

Formation of Striations

An X-ray survey of new striations that form in the crystal independently of those in the seed has given the following results. A series of back-reflection pictures was taken along one pair of striations in line with the sloping shoulder. These striations were not visible for a distance of about one centimeter from the sloping shoulder. The X-ray pictures showed that, in the region where the striations were well developed, there was a fixed difference of orientation between them. As the shoulder was approached and the striations became less distinct and disappeared, the orientation difference decreased until the two spots could scarcely be resolved. This is shown in Figs. 5 to 10.

Although this effect has not yet been fully investigated, it leads to some promising possibilities. Qualitatively, it has been observed that the lower the rate of growth, the greater is the distance from the sloping shoulder to the point where the new striations are visually observable. For sufficiently high rates the new striations form right at the sloping edge. Thus, it would seem that, when new striations must form, as is the case when the geometry of the arrangement makes the crystal wider than the seed from which it is grown, they require an incubation period to develop to their maximum extent as dictated by the conditions of growth.

Discussion

Although the experimental data available are perhaps insufficient to warrant a very exact theoretical analysis, they do suggest a possible explanation of the origin of this phenomenon. Throughout this paper, continual reference has been made to the production of single crystals whereas the resulting product has always consisted of members of differing orientation. However, inasmuch as the orientations of all the members of a specimen are such that they have a common direction and may be brought into coincidence with that of any other member by a rotation of a very few degrees, it seems justifiable to refer to the specimen as a single crystal containing irregularities that result in what could be described as a limited fiber structure.

The geometrically most convenient way that has been suggested of describing the boundary that separates regions of differing orientation is by means of an array of dislocations. Burgers (2) has discussed the dislocation arrays that will describe the boundary that separates regions between which any arbitrary difference of orientation exists. In the present case of striations, the boundary could be described by an array of edge type dislocations as suggested by Burgers or an alternative array of edge dislocations suggested by Shockley and Read (14). It should be borne in mind that the dislocation models that have been proposed and analyzed are strictly applicable only to simple cubic lattices, whereas tin is based on a rather complex body centered tetragonal lattice. In spite of this, it is felt that, to a first approximation at least, a consideration of striation boundaries in terms of simple edge type dislocation arrays would be very profitable.

If this picture of striation boundaries has any basis in fact, there still remains the question of the reason for the formation of dislocations in the first place.

When a solid crystal is heated, it will acquire a concentration of vacant lattice sites, which, according to Frank (8) will be of the order of its vapor concentration for simple molecular crystals and considerably larger for metallic crystals. This concentration will be greatest at the melting point. Calculations of Huntington and Seitz (11) have shown that for copper at the melting point this concentration may be as high as 1 in 10^5 . As the metal crystal is cooled from its melting point, the concentration of vacancies will decrease. Nabarro (12) has suggested a mechanism for the elimination of vacant lattice sites in which the number of vacancies is reduced by their migration to and escape from the surface of the crystal. However, the mobility of vacant lattice sites is considered to be too small for such a mechanism to be operative in any but the very smallest perfect crystals.

A second mechanism that has been suggested by Frank (8) and Seitz (13) seems to be much more probable. As the temperature is lowered from the melting point, the vacancies, lacking other suitable sites to eliminate themselves, will form into aggregates. Both Frank (8) and Crussard (6) have indi-

cated that the most stable form for such an aggregate is in the shape of a flat disk. Condensation of vacant lattice sites into flat disks would continue so that, as the temperature decreases, the disks would grow in size. When a critical size is reached, collapse of the lattice across the disks occurs, producing a general ring dislocation.

This would be the situation if the whole of the solid metal specimen were kept at the same temperature during cooling. In the method of growing crystals from the melt used in this work, the circumstances are quite different. Here, the solid in contact with the melt at the interface is at the melting point and a temperature gradient exists from this interface back into the solid in a direction parallel to the specimen axis. The magnitude of this gradient along with the rate of advance of the interface will govern the manner in which the vacant lattice sites will condense and the subsequent collapse of the disks. The effect of this temperature gradient is that the disks of vacant lattice sites would not all grow and collapse throughout the whole volume of the specimen at the same time, which would be the situation if no gradient existed. Instead, in the solid immediately behind the interface, the disks would be formed and exist as such. A small distance beyond the point where, by virtue of the temperature gradient, the temperature is considerably lower, the disks would have collapsed forming pairs of edge type dislocations. As the interface advances, more vacant lattice sites would condense on the disks at the interface, and those parts of the disks which have now fallen to a lower temperature will collapse, continuing the propagation of the pairs of edge type dislocations.

At the relatively high temperatures immediately behind the interface where these dislocation pairs are formed, the dislocations would be fairly mobile. As Cottrell has pointed out (5), these dislocations would have a tendency to form into transition surfaces consisting of arrays of dislocations of one sign. Cahn (3) has utilized this mechanism in describing the "polygonization" of single crystal specimens that have undergone small amounts of plastic deformation followed by annealing. This would result in just what is observed; i.e., instead of a single crystal, one obtains a specimen which is divided into parallel regions with small differences of orientation between neighboring regions. Also, because of the equality of production of dislocations of opposite sign, one would expect, as is in fact the case, that taken over a sufficiently large number of bands, the rotations in one direction would balance those in the opposite direction.

On this general basis, the effects of changing the rate of growth of the crystal and the orientation of the crystal relative to the direction of longitudinal heat flow may be considered.

The size to which the disks of vacant lattice sites will grow depends on both the temperature and the time for which a given temperature is maintained. The temperature will control the degree of mobility of the vacancies that permits their agglomeration into disks, and the time limits the interval in which the mobility of the vacancies is sufficiently high to make aggregation possible.

Therefore, with a given temperature gradient in the solid, the smaller the period of time allowed for the aggregation of vacancies, the smaller will be the size of the vacancy disks. Consequently, a higher rate of advance of the interface will result in collapse of the lattice occurring across vacancy disks that are smaller than for lower rates of growth.

In addition, the higher rates of growth correspond to shorter times in which the dislocations can assemble into transition surfaces. Thus, these transition surfaces will necessarily be closer together. This produces narrower striations which is in agreement with observation.

When a crystal is grown from a seed in which this mechanism has already produced a striated structure, that part of the crystal that is in line with the seed will grow with a propagation of the striated structure into the crystal. In that part of the crystal not in line with the seed, a new structure must be initiated. On the basis of the mechanism described, one would expect to find an incubation period for the establishment of a stable structure with maximum orientation difference between striations. This too is in agreement with observation.

If, as is very probable, there exist preferred crystallographic planes into which vacancies condense, it would be expected that the actual manner of their condensation would depend on the orientation of these planes relative to the direction of longitudinal heat flow. Thus, the formation of striations would be in a direction that is a compromise between crystallographic and thermal conditions. It has been shown that, when the $[110]$ direction is parallel to the longitudinal heat flow, the striations form parallel to this direction. However, when the $[110]$ direction makes some angle with the specimen axis, the direction of the striations is a compromise between the direction of longitudinal heat flow and the $[110]$ direction, being parallel to the former for low rates of growth and tending closer to the latter at high rates. This would indicate that the vacant lattice sites prefer to condense in planes closely associated with the $\{110\}$ planes.

In conclusion, it should be again emphasized that the theory suggested in the preceding section is, at best, only very qualitative. It does, however, seem to be borne out fairly well by the experimental evidence that is thus far available, and is in itself completely self-consistent.

Summary

In single crystals of tin grown from the melt, a type of lattice imperfection causes the crystal to be partitioned into striations. These striations, which are present throughout the whole volume of the crystal, are such that the lattice in one striation is disoriented from that of its neighbor by a rotation of from $1/4^\circ$ to 5° about the specimen axis. A statistical fluctuation exists in the measured properties of these striations, such as their widths, or their direction of formation. In spite of this, regular variations of these properties have been noted

as functions of the rate of growth and the orientation of the crystal relative to the direction of longitudinal heat flow.

A mechanism is proposed to account for the formation of the striations. This is based on the condensation of vacant lattice sites into disks, and the subsequent collapse of these disks, forming pairs of edge dislocations. These dislocations then migrate into transition surfaces of stable arrays of dislocations of one sign, producing the striation boundaries.

Observations indicate that the $\{110\}$ planes are preferred for the formation of vacancy disks.

Acknowledgments

The authors wish to thank Dr. Ursula Martius of this Department for helpful discussions in this work. The authors also acknowledge the financial assistance of the Research Council of Ontario for scholarships under which this work was carried out, and the National Research Council of Canada for a summer grant. The tin was bought from funds supplied by the School of Engineering Research, University of Toronto.

References

1. BUEGER, M. J. *Z. Krist.* 89: 195. 1934.
2. BURGERS, J. M. *Proc. Phys. Soc. (London)*, 52: 23. 1940.
3. CAHN, R. W. *Proc. Phys. Soc. A*, 63: 323. 1950.
4. CHALMERS, BRUCE. *Proc. Roy. Soc. (London) A*, 175: 100. 1940.
5. COTTRELL, H. H. *Progress in metal physics. Edited by B. Chalmers. Vol. 1. Interscience Publishers Inc., New York.* 1949.
6. CRUSSARD, C. *Métaux & Corrosion*, 25: 203. 1950.
7. DARWIN, C. G. *Phil. Mag.* 27: 315. 1914.
8. FRANK, F. C. *Symposium on Plastic Deformation. Mellon Institute, Pittsburgh. May, 1950.*
9. GRENINGER, A. B. *Am. Inst. Mining Met. Engrs.* 117: 75. 1935.
10. GRENINGER, A. B. *Am. Inst. Mining Met. Engrs.* 117: 61. 1935.
11. HUNTINGTON, H. B. and SEITZ, F. *Phys. Rev.* 61: 315. 1942.
12. NABARRO, F. R. N. *Phys. Soc. (London). Report of Conference on Strength of Solids, p. 75.* 1948.
13. SEITZ, F. *Phys. Rev.* 79: 890. 1950.
14. SHOCKLEY, W. and READ, W. T. *Phys. Rev.* 75: 692. 1949.

DENDRITIC GROWTH IN LEAD¹

BY F. WEINBERG² AND BRUCE CHALMERS³

Abstract

Dendritic growth of large single crystals of lead is revealed by decanting the remaining liquid at any required stage of the process of solidification. The direction of dendritic growth was determined as a function of crystallographic orientation and direction of solidification. It was found that all the dendritic rows in a single crystal were not quite parallel, owing to a crystal substructure. An explanation is advanced for the presence of the substructure in terms of dendritic growth based on their observed properties. Temperature measurements during solidification showed that the region of the liquid where dendrites were formed was supercooled by several degrees. No supercooling was observed in the portions not showing dendritic growth.

Introduction

Although in recent years considerable attention has been paid to the nature of crystal growth, very little has been reported concerning metal crystals grown from the melt, particularly for the high rates of growth at which dendrites have been observed. This is probably due to the opacity of the melt which prevents the direct observation of crystal growth, possible with some organic materials, and to the higher temperatures involved with metals. The authors have therefore examined the dendritic growth of large single crystals of lead, and will present some of the observations. No attempt is made in this paper to present any theory to explain the formation of dendrites during growth, nor will existing theories be discussed.

Cross sections of dendritic structures have been observed for many years on polished and etched sections of castings and are familiar to most metallurgists. The structure consists of skeletons of solid metal, formed during solidification, which grow into the liquid either ahead of the general solid-liquid interface or from a solid particle in the liquid. For cubic metals, their shape is usually described as that of Fig. 1, adapted from the diagram by Carpenter and Robertson (1). This structure has recently been questioned by Northcott (3) in his examination of copper alloy castings.

From the figure, the dendrite is seen to consist of a primary horizontal stalk from which small secondary stalks grow periodically in the horizontal and vertical planes. From these secondary stalks, smaller tertiary branches grow parallel and perpendicular to the primary stalk and so on. It is believed that, under proper growth conditions, a fine primary stalk initially forms from the solid. Secondary branches then grow as the stalk becomes longer and thicker, followed by tertiary branches, the entire skeleton becoming progressively

¹ Manuscript received February 14, 1951.

Contribution from the Department of Metallurgical Engineering, University of Toronto, Toronto, Ont.

² Graduate Student, Department of Metallurgical Engineering, University of Toronto.

³ Professor of Physical Metallurgy, Department of Metallurgical Engineering, University of Toronto.

larger and thicker as growth proceeds. The direction of growth of the primary stalk appears to depend on the crystallographic orientation of the parent metal, Northcott and Thomas (4) having shown for copper alloys that dendritic markings on a polished surface were outcrops of $\{100\}$ planes.

Since it is to be expected that dendritic growth may be explained in terms of atomic arrangements of the solid adjacent to the liquid, or in terms of heat flow, both of which are markedly influenced by impurities, only very high purity (99.999%) lead was used. It was also apparent that, if there was regularity in the dendritic pattern, it should be more readily observable with large crystals, and therefore relatively large crystals were grown and examined.

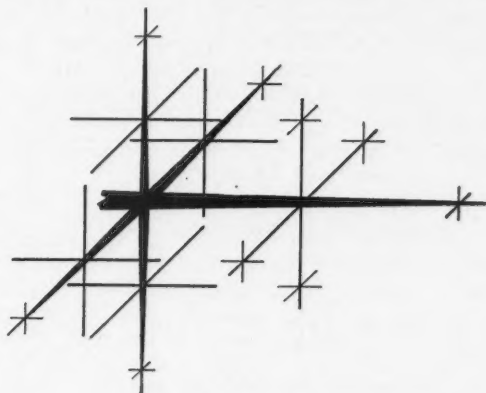


FIG. 1. *Dendrite skeleton.*

A. Papapetrou (5) and others have advanced an explanation of the dendritic growth of nonmetallic crystals from solutions in terms of diffusion phenomena, the essential condition being the occurrence of supersaturation of the liquid. In an analogous manner, they believe that dendritic growth for metallic crystals may be associated with supercooling of the melt.

A series of cooling curves was therefore obtained for various stages of growth of the crystals in order to obtain further experimental evidence relating supercooling to dendritic growth.

Experimental Procedure

Method of Growth

Because of the high purity of the metal used, the crystals could not be polished and etched to reveal their dendritic structure, since this method implies the presence of a second constituent or impurity in the metal along the dendrite boundaries. Further, it confines all observations to the plane of polish, showing only a section of a three dimensional dendrite. An alternative method was

therefore adopted. This consists of pouring off the remaining liquid at the desired stage of growth. The solid that has formed remains in the boat and it can be examined at leisure.

The crystals were grown in the manner described by Chalmers (2) in which the "seed crystal", a single crystal of the desired orientation, is placed in a horizontal graphite boat containing some polycrystalline metal. This metal, and part of the seed adjacent to it, is then melted by a small furnace surrounding the boat, and the liquid is agitated to make good contact between liquid and seed. The furnace is then slowly moved away from the seed along tracks parallel to the boat. The resulting crystal is generally a single crystal having the same orientation as the seed and grown at a rate governed by the rate at which the furnace moved during solidification of the metal. A seed crystal of a different orientation is grown as above, but the original seed is tilted and rotated by the required amount with respect to the boat axis. The boat consists of a strip of 1/2-in. thick graphite in which a trough 3.4 cm. wide and 1.0 cm. deep has been milled, leaving side walls 0.2 cm. thick. The furnace is mounted on wheels and consists of a dozen turns of nichrome wire surrounded by fire brick plates to reduce heat losses.

If the liquid is agitated during growth, a ripple mark is frozen in along the top edge of the solid-liquid interface. By agitating at equal time intervals, the approximate rate of growth of the crystal may be determined from subsequent measurements of the position of the corresponding ripple marks. The crystals were grown in air, and the oxide which formed on the liquid surface was skimmed off with a flattened glass rod.

During the growth of a number of crystals at rates controlled by uniform movement of the furnace, the furnace was suddenly pulled away and the end of the boat dropped, allowing the remaining liquid to run down the boat away from the solid metal. It was found that at slow rates of growth (less than 1 cm. per minute) the solid-liquid interface was essentially plane and perpendicular to the direction of growth. At higher rates, the interface became progressively more concave, owing to the higher thermal conductivity of the graphite. In no case, even up to rates of 2.5 cm. per min., were dendrites observed.

However, if the furnace was completely pulled away from the boat after solidification had started, and the crystal allowed to freeze at a rate governed only by the rate at which heat was lost to the surrounding atmosphere, it was found that large dendrites could be obtained. This procedure was therefore adopted for this investigation, and all the crystals illustrated were grown in this manner. The rate of growth measured on the top surface now appeared to accelerate roughly exponentially until the first dendrite peak appeared at the surface. Following this, the rate of growth could not be defined since the nominally "liquid" metal consisted partly of solid dendrites and partly of liquid until solidification was complete.

The seed crystals used were usually 3 cm. long and were the full width of the boat, to ensure symmetrical heat flow in the horizontal plane.

Bicrystals were grown in essentially the same manner as that of the single crystals. In the place of the one wide seed crystal, two narrow seeds of the desired orientations were placed in the boat side by side and separated by a short graphite insert. A crystal was then grown at a uniform rate simultaneously from both small seeds to a point well past the end of the insert. The result was usually a bicrystal which was the full width of the boat. This was subsequently used as the seed for crystals grown in the usual manner.

Measurement of Temperature

A small thermocouple was inserted horizontally into the molten lead through holes made in the side of the boat. The temperature of the metal, as indicated by the thermocouple, was then recorded as the crystal grew.

The thermocouple was made by silver soldering copper and constantan wires having diameters of 0.0048 in. The copper wire had a glass fiber wrapping which satisfactorily insulated it from the bare constantan wire for the temperature range encountered. The protecting tube surrounding the wires was made of glass and had an outside and inside diameter of 0.9 mm. and 0.5 mm. respectively.

Preliminary temperature measurements with a Leeds and Northrup Type K2 potentiometer showed that it was necessary to immerse the thermocouples at least 18 mm. into the liquid to obtain consistent results at the freezing point of lead. With sufficient immersion, the measured temperatures are believed to be correct to 0.1 degree Centigrade.

After calibration, the thermocouple was placed 1.5 mm. above and parallel to the bottom of the graphite boat with the junction 2.5 cm. from one side wall. Molten metal was then added to the boat and a crystal 5 mm. thick was grown in the usual manner from a seed at the end of the boat. Most of the resultant e.m.f. of the immersed couple during growth was balanced with the potentiometer kept at a fixed position. The remainder was recorded on a Leeds and Northrup Speedomax recorder having a full scale sensitivity of 5 mv., giving temperature measurements believed correct to 0.2 degree Centigrade, and temperature differences taken from one curve, correct to 0.1 degree Centigrade.

A series of cooling curves was then obtained for one crystal which was regrown a number of times. In each case, the point at which solidification started, with respect to the fixed couple, was changed, effectively changing the position of the couple in the crystal. The necessary amount of metal was added to keep the total amount of metal solidifying in each case constant.

An attempt was made to grow crystals thicker than 5 mm. in order to decrease the ratio of thermocouple thickness to specimen thickness; this would reduce any disturbance in the system due to the presence of the thermocouple. However, the attempt was abandoned, since it was found that stray crystals tended to form in the thicker crystals when they were reheated, even with very careful handling. The possibility of introducing a number of thermocouples at various fixed positions in the crystal was also considered. It was not adopted, since the presence of a number of couples could appreciably change the normal thermal conditions during crystal growth.

Experimental Observations and Discussion

Dendrite Structure and Manner of Growth

A typical example of the dendritic structure obtained is shown in Fig. 2, which is a photograph of a decanted single crystal of lead viewed from above and to one side. The crystallographic orientation is indicated by the three arrows, each arrow being parallel to a $\langle 100 \rangle$ crystallographic axis or cube edge, since lead has a face-centered cubic structure. Growth proceeded from the left from a point outside the photograph.

The dendritic structure is seen to consist of a series of parallel "dendrite rows", each row corresponding to the upper half of the dendrite skeleton drawn in Fig. 1. Taking the primary stalk of the skeleton as being in the horizontal plane at the bottom of the crystal, the vertical secondary and horizontal tertiary stalks in one row are clearly seen, growing in a direction corresponding to the diagram. The tertiary branches growing along the surface of the liquid branch even further as shown in Fig. 3, which is a photograph showing the details of part of the top surface of a vertical stalk. In some cases, the larger vertical stalks have been disturbed in the decanting process and are tilting over, the smaller ones remaining perpendicular to the primary stalk.

In Fig. 1, the secondary branches are shown progressively decreasing in length and thickness as one proceeds from the center to the outer ends of the primary stalk; this, however, is not the case for the actual dendrites shown in Fig. 2. On examining the vertical stalks, for example, it is apparent that a relatively large number of stalks begin to grow, but only a few reach the top surface where the vertical growth is arrested. Nor do the vertical stalks which do develop grow at equal intervals as shown in Fig. 1. It would also be expected that, since the secondary stalks grow from the bottom of the boat, the tertiary branches should be larger and thicker the closer they are to the bottom; this is not the case here for the larger stalks.

The manner in which the dendrites grow is indicated by the three crystals shown in Fig. 4. These crystals were grown under similar conditions, but were decanted at 50, 52, and 55 sec., for crystals (a), (b), and (c) respectively, from the instant at which the furnace was removed. The upper edge of the solid-liquid interface of crystal (c) had advanced 5 mm. further from the point at which solidification started than that of (a). Comparing the three crystals, it is seen that the dendritic rows of crystal (a) are shorter and thinner than those of the other two. Their secondary branches are very small, showing no further branching, and the spacing between rows is smaller than that of the other two. The photograph does not show some of the smaller rows for this crystal. The same to a lesser degree may be said of the dendrite size and spacing of (b) with respect to (c). The manner in which the dendrites grow therefore appears to agree generally with the description given by Carpenter and Robertson, the primary stalk growing first, followed by the secondary and tertiary stalks, some of which progressively increase in size. As noted before, this is not true for the vertical stalks which reach the top surface, since their growth is arrested

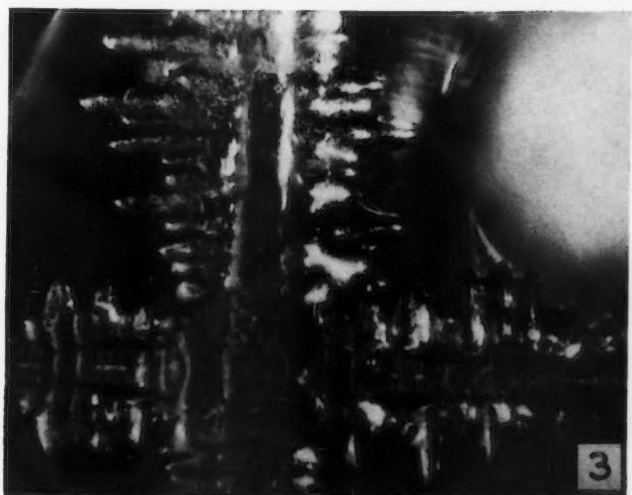
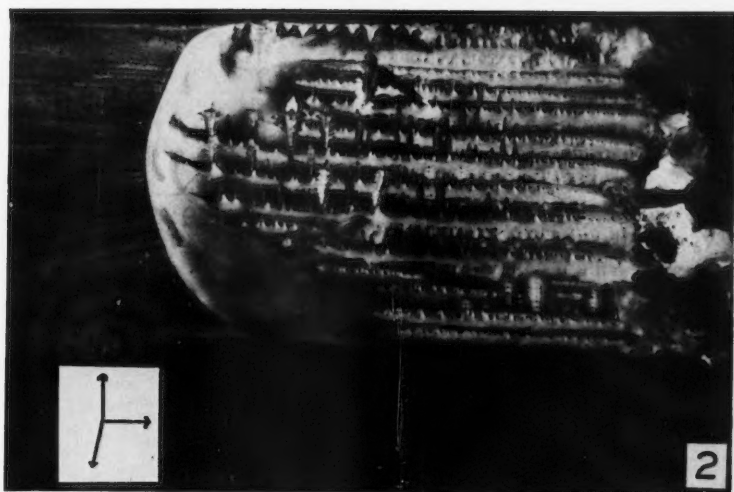


FIG. 2. A decanted lead single crystal viewed obliquely from above. The arrows are parallel to the crystallographic axes.

FIG. 3. Details of the top branches of a vertical dendrite stalk.

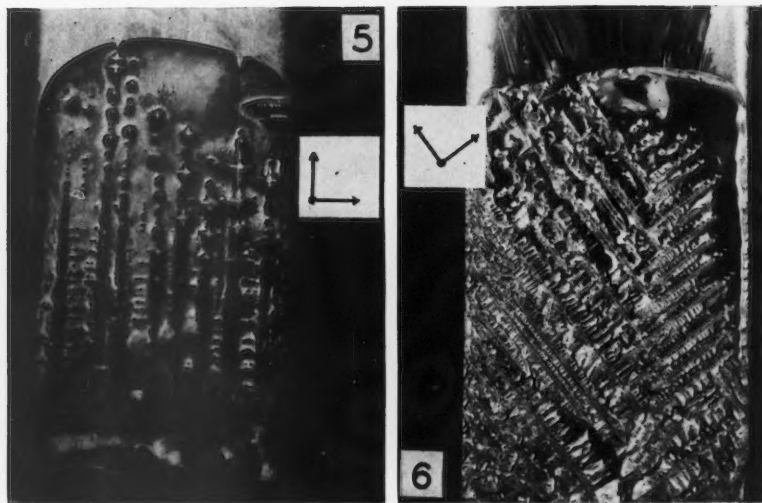
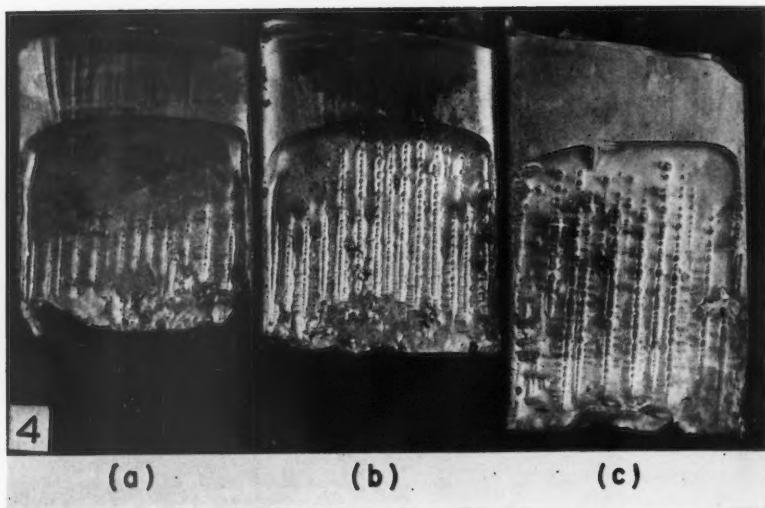


FIG. 4. (a), (b), and (c). Crystals decanted at progressively later stages of solidification.

FIGS. 5 AND 6. Plan view of the dendrite pattern for two crystals grown under similar conditions. The arrows are parallel to their crystallographic axes.

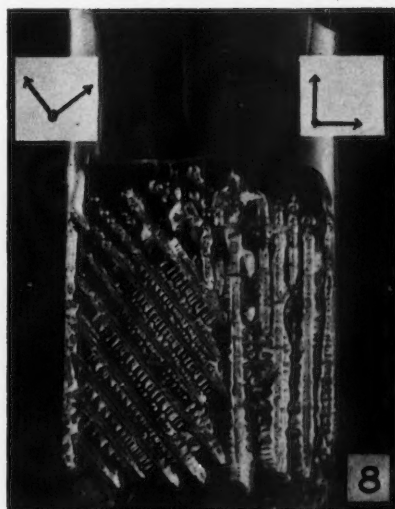
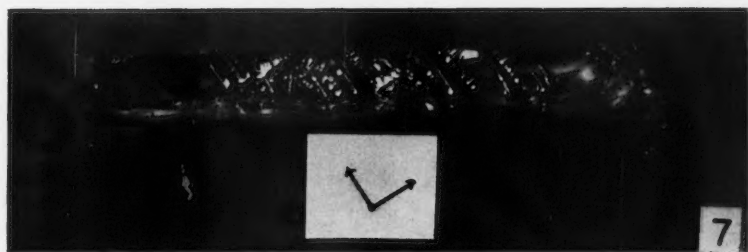


FIG. 7. End view of a decanted crystal orientated as shown.

FIG. 8. Dendritic pattern of a bicrystal whose components are orientated similarly to those of Figs. 5 and 6.

FIG. 9. Bottom view of a crystal showing the transition from striations at the top to jagged boundary zones.

PLATE IV



FIGS. 10 AND 11. Top and bottom views of the same crystal, both showing the same zones.
 FIG. 12. Crystal showing a fern type dendrite which had been growing along the top surface of the liquid but which fell to the bottom on decanting.
 FIG. 13. Crystal showing convex part of the interface under which the oxide between the liquid metal and boat was removed

by the surface and their branches grow much larger at the top. At the same time that the dendrites are growing, the general concave solid-liquid interface sweeps along covering the rows initially formed.

Direction of Growth

The relation between crystallographic orientation and direction of dendrite growth is further shown in Fig. 5, which is a plan view of a crystal having the same orientation as that of Fig. 2. The mutually perpendicular directions of growth of primary, secondary, and tertiary branches is clearly seen, although two vertical stalks have fallen over in the decanting process. The crystal shown in Fig. 6 was grown under similar conditions to those above except that the [110] direction is now parallel to the boat axis instead of the [100] direction. The dendrite rows and smaller branches grow at 45° to the general direction of growth, but still parallel to the cube edge of the crystal lattice. To ensure that growth conditions were the same for both crystals, a bicrystal was grown and the result is shown in Fig. 7. The crystal on the right has the same orientation as that of Fig. 5, that on the left the same as that of Fig. 6. It therefore appears that the direction of dendrite growth depends only on the crystallographic orientation of the parent crystal and not on the direction of solidification.

Crystals of other orientations were grown and decanted, and the results confirmed the above statement. Fig. 8 shows an end view of a crystal whose orientation is such that one axis lies along the longitudinal crystal axis, the others being at approximately 45° to the horizontal plane. As expected, the secondary stalks now grow correspondingly at 45° to the horizontal plane. For the cases in which a cube edge was not parallel to the longitudinal crystal axis, as for example is the case for the [111] direction parallel to this axis, dendrite rows were not observed, but stalks were observed near the completely solidified crystal, projecting in directions corresponding to the cube edges of the crystal. It should be noted that crystal (c) of Fig. 4 has not quite the same orientation as that of (a) and (b). This accounts for the small difference in dendritic row direction between it and the other two.

In all cases observed, the dendrite direction appeared to correspond to within two degrees of the crystallographic direction, this value being the limiting accuracy for the experimental procedure adopted. However, when crystals similar to that of Fig. 2 were examined more closely, it was found that all the dendrite rows were not exactly parallel, some rows diverging from their neighbors by as much as three degrees. At the same time, when the crystals were deeply etched in a 10% nitric acid solution, it was found that all the crystals did not have the same orientation as seen from the different optical reflective properties of each section and verified by back-reflection X-ray photographs. This phenomenon will be discussed more fully in the next section.

To determine the effect of orientation differences on direction of dendrite growth, six bicrystals were grown in which one component of the bicrystal was related to the other by a simple rotation about the vertical axis, as in Fig. 7. Measurements were then made of the position of the dendrite rows with a low

powered microscope and travelling stage, and a series of X-ray photographs were taken traversing the crystal. The divergence of the dendrite rows in two zones, one in each component of the bicrystal, could then be related to the orientation difference of the two zones. The results are listed in Table I.

TABLE I

Bicrystal	Dendritic row divergence	Orientation difference
<i>A</i>	4° 35'	5°
<i>B</i>	6 15	6
<i>C</i>	16 5	16
<i>D</i>	21 20	22
<i>E</i>	37 50	37
<i>F</i>	54 10	55

The accuracy of measurement of the direction of a dendrite row is limited by the round tips of the vertical stalks which are lined up on the cross hairs of the microscope. However, in these experiments, this error is less than the probable error in the measurement of the relative orientation of the zones, estimated at one degree. This is due to the size of the Laue spot on the film, the X-ray spot falling on the specimen being 0.7 mm. in diameter.

It may, therefore, be concluded that, within experimental error, the direction of dendritic growth is uniquely defined by the crystallographic orientation of the dendrites and is independent of the direction of heat flow. The presence of zones of slightly different orientations could account for the fact previously noted that the dendrite rows of a single crystal are not quite parallel.

Crystal Substructures at High Rates of Growth

At slow rates of growth "striation" substructures have been described for tin single crystals by Teghtsoonian and Chalmers (6) who have also referred to the existence of similar structures in lead. They report that striations appearing on the top surface of a crystal consist of a series of bands parallel to the longitudinal axis of the crystal, with an average width of about a millimeter. The crystallographic orientation of the striations with respect to any one reference striation is defined by a small simple rotation about the longitudinal axis of the crystal. At higher rates of growth, the average width of the striations decreases and they converge, since they grow perpendicular to the solid-liquid interface which becomes concave for crystals grown in the present manner. They can be seen in Fig. 2, the striations appearing on the completely solidified portion of the crystal on the left-hand side of the photograph.

When crystals of lead showing dendritic growth were deeply etched and examined, striations were observed on the bottom of the slowly grown seed as well as the top. However, as the rate of solidification increased, the appearance and nature of the striations changed as shown in Fig. 9. This is the bottom view of an etched crystal similar to that of Fig. 2, in which some of the striations in the seed crystal appear at the top of the photograph. It can be seen

that a few striations widen considerably at the expense of the others. Further, the boundaries between them are no longer straight lines but become jagged and do not remain parallel on the average to the longitudinal axis of the crystal. An X-ray examination of these jagged boundary zones shows that each zone is of uniform orientation, but neighboring zones appear to differ in relative orientation by small amounts, up to three degrees. There appears to be no common axis of rotation relating the zones in any one crystal. The average direction of the jagged boundaries tends to be parallel to the dendrite rows.

Looking at the top surface of the crystal, jagged irregular boundaries corresponding to those observed on the bottom are seen in the decanted portion of the crystal. This is shown in Figs. 10 and 11. Fig. 10 is the plan view of the top surface of a crystal illuminated to bring out some of the zones. Four zones can be seen in the lower half of the crystal. Fig. 11 is a photograph of the bottom surface of the same crystal, showing the same zones. As shown in Fig. 10, the jagged boundaries run between dendrite rows appearing to be a compromise between the horizontal side arms projecting from the primary stalks of the rows on either side of the boundary. In all of the specimens examined, no dendrite row ever crossed one of these boundaries, although in some cases rows terminated when they encountered one. Also in some cases where large dendrites grew from the lip from the side of the crystal toward the center, before the dendrite rows had grown that far, the boundary turned sharply and followed the direction of these lateral dendrites. The boundaries in the decanted portion could be clearly distinguished up to very nearly the upper edge of the concave solid-liquid interface of the crystal where the regular striations appear. In one case, the crystal was electrolytically etched in a weak solution of nitric acid. This resulted in the striations' receding approximately 14 mm. toward the seed after 0.27 mm. of material had been removed from the top surface, bringing out zones which were continuations of those observed in the decanted portion. Further etching did not materially affect the end position of the striations.

Jagged boundaries were also observed on the bottom surface of bicrystals. The degree of irregularity does not appear to be affected by the difference in orientation of the crystals on each side of the boundary, but appears to be more irregular in the portions of the crystal that solidify last where the rate of growth is probably greatest.

A tentative explanation of the development of these jagged boundary zones is advanced in terms of dendritic growth. At rates of growth greater than about one centimeter per minute a lip is observed on the decanted surface, projecting above the bottom of the boat at the foot of the nearly plane vertical solid-liquid interface, indicating that the part of the liquid adjacent to the graphite boat is slightly cooler than the remainder. Therefore, it is possible that very small flat "fern type" dendrites could grow along the bottom of the boat taking up the orientation of the striations from which they originate. However, as a result of the slight differences in thermal contact between the liquid and the graphite boat, some dendrites may grow ahead of others, the former being the

basis of the zones. Once this condition has been reached, the lateral branches of the leading dendrites will grow in front of the slower dendrites preventing their further growth. It is also possible at this stage that convection currents in the liquid adjacent to the solid, or vibration, could bend the fine dendrites, accounting for the small orientation differences of neighboring zones. When the lateral branches of the remaining dendrites meet, one would expect a jagged boundary as a compromise between the two, as is observed, depending only on the outline of the dendrites and not on their relative orientations.

As noted before, striations grow perpendicularly to the solid-liquid interface; therefore, it would be expected that those growing along the bottom of the crystal would tend to swing towards the top in order that they remain perpendicular to the lip surface. If new striations did not form, the metal below the striations would then take up the orientation of the solid defined by the small dendrites. Therefore, since the lip becomes progressively larger as growth accelerates, the zones forming from the small dendrites along the bottom would be expected to become thicker until by the time the decanted portion of the crystal is reached, they would extend through the entire thickness of the crystal. Since the very top of the solid-liquid interface remains essentially vertical, as shown in Fig. 2, the striations would be expected to extend much farther along the crystal at the top than at the bottom. They should disappear near the decanted portion when a layer of the top surface is removed by etching, as was observed.

The fern type dendrites referred to have been observed in a number of cases growing from the solid along the top of the liquid exposed to the air. As the name implies, they are flat with closely spaced side branches and often project several centimeters ahead of the interface into the liquid. An example is shown in Fig. 12 in which the fern type dendrite has fallen to the bottom when the liquid under it was poured away.

The effect of thermal contact between the liquid and the graphite boat on rate of growth is shown markedly in Fig. 13. Normally, when metal is added to the boat and melted, a layer of oxide forms both on the top surface of the liquid and on the bottom surface between the liquid and graphite boat. The top layer of oxide is removed, but the bottom layer usually remains, resulting in poorer thermal contact between liquid and boat. In the crystal in Fig. 13, the oxide was removed along a strip running down the center of the crystal, by rubbing the boat surface under the liquid with a flattened glass rod. It is seen that the rate of solidification above the strip was greater than that of the remainder, as shown by the convex portion of the general concave solid-liquid interface. If removal of the oxide affects the rate of solidification to the extent shown, then it seems feasible that slight differences of temperature could exist along the boat surface to produce the advance of several tiny dendrites ahead of their neighbors.

Northcott (3) notes in his examination of a dendritic structure brought out by polishing and etching a section of a copper alloy casting that irregularities

of the structure were present on what appeared to be a single crystal. X-Ray examination showed small differences in orientation between these regions. It appears that this irregularity could be due to the presence of zones similar to those described above.

Supercooling

Some of the cooling curves obtained in the manner previously described are shown in Fig. 14. Their corresponding position in the crystal is given by d , the distance between the thermocouple and the point at which solidification started, in Table II. This table includes the supercooling indicated by the

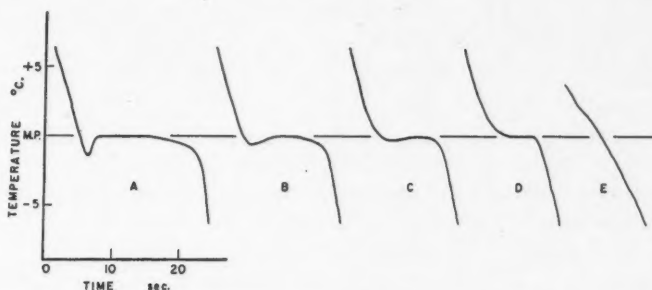


FIG. 14. Cooling curves for various positions in the crystal.

TABLE II

Curve	d , cm.	Supercooling, °C.	Dendrite appearance
<i>A</i>	6.5 5.5 4.5	1.2	Well developed—as in the decanted portion of Fig. 2
<i>B</i>	3.5	0.5	Partly developed—as in Fig. 4 (<i>b</i>)
<i>C</i>	2.5	0.2	Beginning—as in Fig. 4 (<i>a</i>)
<i>D</i>	1.5	0.0	None
<i>E</i>	0.5	0.0	None

curves and the dendritic pattern observed in the region surrounding the thermocouple for crystals grown under similar conditions.

From the last two columns of Table II, it appears that supercooling is a necessary condition for dendritic growth in high purity lead. It also appears that the development of the dendrite branches is related to the degree of supercooling, since for thermocouple positions giving the same supercooling (Curve *A*) the dendrite development is generally the same; for less supercooling (Curves *B* and *C*) the development is correspondingly smaller.

It has previously been noted that dendrites were not observed when crystals were grown at uniform rates. It would therefore be expected that under these

growth conditions the liquid ahead of the solid-liquid interface should not be supercooled. This was found to be so, within the sensitivity of the measuring apparatus.

Summary

For high purity lead, it is concluded that a dendrite skeleton forms from a small single stalk projecting into the liquid ahead of the general solid-liquid interface. The secondary branches of the skeleton grow from and perpendicular to the primary stalk as it grows larger and thicker. Some of these secondary branches become larger as growth proceeds and develop tertiary branches which, in turn, branch again and so on.

All the stalks grow parallel to $\langle 100 \rangle$ directions defined by the crystallographic orientation of the dendrite. Their directions are not influenced by the temperature gradients present in the liquid in which they grow.

At high rates of growth, zones of similar crystallographic orientation are observed in the crystal, each zone having a small difference in orientation with respect to its neighbor. Because of this orientation difference, all the dendrite rows in a single crystal are not exactly parallel, since their direction is defined by the orientation of the zone in which they grow.

It is believed that the orientation of the zones is determined by the initial tiny dendrites which develop during growth and which in some cases have been disturbed by convection currents and vibrations in the liquid surrounding them. The zones are characterized by jagged irregular boundaries which appear to be a result of a compromise between the side branches of the two dendrites growing on either side of the boundary. They are not observed in crystals grown at slow rates of growth.

From temperature measurements, it appears that supercooling is a necessary condition for the dendritic growth of high purity lead, and that the size and spacing of the dendrite stalks is a function of the degree of supercooling of the liquid in which they form.

Acknowledgments

We wish to thank the Consolidated Mining and Smelting Co. for kindly supplying the special high purity lead. We are also indebted to Prof. K. B. Jackson of the Applied Physics Department for assistance with the photography, and to the School of Engineering Research and the National Research Council of Canada for financial aid.

References

1. CARPENTER, H. and ROBERTSON, J. M. *Metals*. Vol. 1. Oxford University Press. 1939.
2. CHALMERS, B. *Proc. Roy. Soc. (London)*, A, 196:64. 1949.
3. NORTHCOTT, L. *J. Inst. Metals*, 72: 283. 1946.
4. NORTHCOTT, L. and THOMAS, D. E. *J. Inst. Metals*, 65: 205. 1939.
5. PAPAPETROU, A. *Z. Krist.* 92: 89. 1935.
6. TEGHTSOONIAN, E. and CHALMERS, B. *Can. J. Phys.* 29: 370. 1951.

THEORY OF MULTIPOLE RADIATIONS¹

By P. R. WALLACE

Abstract

We develop a systematic analysis of the radiation from a given oscillating system of charges and currents, without any approximations. Using a simple vector identity, the vector potential is separated into irrotational and solenoidal parts. The field may be expressed in terms of the latter alone. A similar vector identity involving the operator $\mathbf{L} = \mathbf{r} \times \text{grad}$ (the rotation operator) permits the separation of the field into parts in which the radial components of the electric and magnetic field, respectively, vanish. The energy flux, energy density, and angular momentum density may in each case be expressed in terms of the angular operators \mathbf{L}, \mathbf{L}^2 . Expansion in the eigenfunctions of these operators, the spherical harmonics, corresponds to the separation into electric and magnetic multipoles of all orders. Introduction of "tensor spherical harmonics" enables us to exhibit these radiations in terms of natural multipoles (derivatives of $1/r$). All calculations are carried out without restriction as to size of radiating system relative to wave length, in the induction as well as the radiation region.

1. Introduction

There exists an extensive literature on the theory of multipole radiation (2-9). Some of this is concerned with specifying the form of the radiation fields for the various radiations, without reference to the source of the radiation. The most elegant treatment is undoubtedly that of Franz (4). Those papers that deal with the radiation from a given system of oscillating charges and currents are in general not free of approximations based on the assumption of a wave length large compared with the dimensions of the radiating system, or are restricted to the distant field only, or to the lower multipole orders.* The present paper represents an attempt to give a general treatment, using only elementary mathematical techniques. The aim is to proceed as far as possible with the analysis of the field merely into "electric" and "magnetic" parts, and to express the energy, angular momentum, etc., of each. Expansion of these quantities in spherical harmonics then corresponds immediately to separation into multipole orders.

2. Representation of the Field by a Vector Potential

We shall suppose that the source of the radiation consists of a charge distribution $\rho(x', y', z', t) = \rho_0(x', y', z')e^{-i\omega t}$ and a current distribution $\mathbf{J}(x', y', z', t) = \mathbf{J}_0(x', y', z')e^{-i\omega t}$. The equation of continuity

$$\text{div } \mathbf{J} + \frac{\partial \rho}{\partial t} = 0$$

¹ Manuscript received April 2, 1951.

^{*} Contribution from the Department of Mathematics, McGill University, Montreal, Que.

^{*} However, Dr. J. Blatt of the University of Illinois has been kind enough to show the author an alternative general treatment, in the spirit of the paper by Franz, which will appear in a forthcoming text book (on nuclear physics) by Weisskopf and Blatt.

^{**} If the time variation is not simple harmonic, the charges and currents must be analyzed into Fourier components; the field is then a superposition of oscillatory components with a distribution of frequencies.

becomes in this case

$$\operatorname{div} \mathbf{J} = i\omega\rho \quad (2.1)$$

In terms of ρ and \mathbf{J} the scalar and vector potentials of the field may be written

$$\begin{pmatrix} \phi \\ \mathbf{A} \end{pmatrix} = \int \begin{pmatrix} \rho(t-R/c) \\ \frac{1}{c} \mathbf{J}(t-R/c) \end{pmatrix} \frac{1}{R} dV' \quad (2.2)$$

where $R = |\mathbf{r} - \mathbf{r}'|$, \mathbf{r} being the point at which the potentials are being evaluated and \mathbf{r}' the variable of integration ranging over the distribution. With the given distributions (2.2) may be written

$$\begin{pmatrix} \phi \\ \mathbf{A} \end{pmatrix} = \int \begin{pmatrix} \rho(\mathbf{r}', t) \\ \frac{1}{c} \mathbf{J}(\mathbf{r}', t) \end{pmatrix} \frac{e^{ikR}}{R} dV', \quad (2.3)$$

where $k = \omega/c = 2\pi/\lambda$, λ being the wave length of the radiation.

Consider now the vector identity

$$(\mathbf{J} \times \nabla') \times \nabla' = -\mathbf{J} \nabla'^2 + (\mathbf{J} \cdot \nabla') \nabla' \quad (2.4)$$

where ∇' is the gradient operator with respect to the primed ("internal") coordinates. Letting this operate on e^{ikR}/R and using the fact that

$$\nabla'^2 \frac{e^{ikR}}{R} + k^2 \frac{e^{ikR}}{R} = 0,$$

we get for \mathbf{A} :

$$\mathbf{A} = \frac{1}{k^2 c} \left[-\int (\mathbf{J} \cdot \nabla') \nabla' \frac{e^{ikR}}{R} dV' + \int (\mathbf{J} \times \nabla') \times \nabla' \frac{e^{ikR}}{R} dV' \right]. \quad (2.5)$$

We may now use the fact that

$$\nabla' \frac{e^{ikR}}{R} = -\nabla \frac{e^{ikR}}{R} \quad (2.6)$$

and integrate the first term by parts to obtain

$$\mathbf{A} = \frac{1}{k^2 c} \left[-\nabla \int \operatorname{div} \mathbf{J} \frac{e^{ikR}}{R} dV' + \nabla \times \left(\nabla \times \int \mathbf{J} \frac{e^{ikR}}{R} dV' \right) \right] \quad (2.7)$$

The two parts of \mathbf{A} are respectively irrotational and solenoidal. The first therefore contributes nothing to the magnetic field. Its contribution to the electric field, $-\frac{1}{c} \frac{\partial \mathbf{A}^{(1)}}{\partial t}$, may be written in terms of the charge density by virtue of the continuity relation (2.1); it is $\nabla \int \rho \frac{e^{ikR}}{R} dV'$ and therefore cancels $-\nabla\phi$. Thus, the whole field may be derived from the vector potential

$$\mathbf{A} = \frac{1}{k^2 c} \nabla \times \left[\nabla \times \int \mathbf{J} \frac{e^{ikR}}{R} dV' \right]^*. \quad (2.8)$$

* If, in addition to currents, there is an oscillating density of magnetization \mathbf{I} , one must add to \mathbf{J} a term $c \operatorname{curl} \mathbf{I}$. This makes it possible to take account of spin contributions in quantum-mechanical problems.

3. Separation into "Electric" and "Magnetic" Radiations

The key to the separation of (2.8) into "electric" and "magnetic" radiations is the introduction of the "rotation operator"

$$\mathbf{L} = \mathbf{r} \times \nabla. \quad (3.1)$$

This operator is the same as the quantum-mechanical angular momentum operator to within a (complex) factor. It may also be envisaged as follows: if $\phi(x, y, z)$ is defined as a function of position, $\omega \cdot \mathbf{L}\phi$ is the rate of change of ϕ as the point x, y, z is rotated with angular velocity ω about the origin. It follows that if ϕ is expressed in polar coordinates and is expanded in spherical harmonics, the radial dependence is unaffected by the operator \mathbf{L} ; i.e. \mathbf{L} is an operator on the angular function only. A useful result, which follows immediately, is the following: if f is a function of the angle θ between the position vectors \mathbf{r} and \mathbf{r}' , then

$$\mathbf{L}'f = \mathbf{r}' \times \nabla'f = -\mathbf{L}f. \quad (3.2)$$

This follows from the fact that a rotation of \mathbf{r}' about any axis induces the same change in θ and therefore in f as an equal and opposite rotation in \mathbf{r} .

The splitting of the vector potential (2.8) now follows by the use of a vector identity similar to (2.4), but with \mathbf{L} -operators replacing the gradients:

$$(\mathbf{J} \times \mathbf{L}) \times \mathbf{L}' = -\mathbf{J}(\mathbf{L} \cdot \mathbf{L}') + (\mathbf{J} \cdot \mathbf{L}')\mathbf{L}. \quad (3.3)$$

If this operates on a function of θ , by virtue of (3.2)

$$\mathbf{L} \cdot \mathbf{L}' = -\mathbf{L}^2,$$

and we may write (3.3) symbolically as

$$\mathbf{J} = \mathbf{L}^{-2}[(\mathbf{J} \times \mathbf{L}) \times \mathbf{L}' - (\mathbf{J} \cdot \mathbf{L}')\mathbf{L}] \quad (3.4)$$

where \mathbf{L}^{-2} is the inverse operator of \mathbf{L}^2 . Let us multiply on the right by e^{ikR}/R . Now introduce a function g such that

$$\mathbf{L}^2 g = \frac{e^{ikR}}{R} - \left\langle \frac{e^{ikR}}{R} \right\rangle_{\text{Av}} = f \quad (3.5)$$

where $\left\langle \right\rangle_{\text{Av}}$ designates the average over-all directions.

Since $\mathbf{L}^2 \mathbf{P}_l(\cos \theta) = -l(l+1) \mathbf{P}_l(\cos \theta)$,

using the known expansion for e^{ikR}/R in spherical harmonics, we find that

$$g = ik \sum_{l=1}^{\infty} \frac{2l+1}{l(l+1)} j_l(kr') h_l^{(1)}(kr) \mathbf{P}_l(\cos \theta), \quad (3.6)$$

$j_l(kr')$ being the "spherical Bessel function" of order l , and $h_l^{(1)}(kr)$ the spherical Hankel function of the first kind (diverging wave).

Now since

$$L \frac{e^{ikR}}{R} = Lf,$$

and

$$L^{-2}Lf = LL^{-2}f = Lg$$

we get, by letting (3.4) operate on e^{ikR}/R and integrating,

$$\int J \frac{e^{ikR}}{R} dV' = \int (J \times L) \times L'g dV' - \int (J \cdot L') Lg dV'. \quad (3.7)$$

Therefore

$$A = A_1 + A_2$$

where

$$A_1 = \frac{1}{k^2 c} \nabla \times [\nabla \times \int (J \times L) \times L'g dV'], \quad (3.8)$$

$$A_2 = -\frac{1}{k^2 c} \nabla \times [\nabla \times \int (J \cdot L') Lg dV']. \quad (3.9)$$

In the second,

$$\nabla \times [\nabla \times L] = \nabla(\nabla \cdot L) - \nabla^2 L = -L\nabla^2,$$

and since $\nabla^2 g = -k^2 g$

the second becomes

$$A_2 = -\frac{1}{c} L \int (J \cdot L') g dV'. \quad (3.10)$$

This, we shall see, is the potential of the *magnetic* multipole radiations, while A_1 gives that of the *electric*.

We shall now transform A_1 into a more convenient form. In Equation (3.8), for A_1 , the last L' may be replaced by $-L$. Then we may write

$$\{\nabla \times [(J \times L) \times L]\}_x = \nabla \cdot (J \times L)_x L + (J \cdot \nabla \times L)L_x$$

etc. Using the commutation rules for the components of L :

$$L_x L_y - L_y L_x = -L_z \text{ etc.} \quad (3.11)$$

and the fact that $\nabla \cdot L = 0$, this becomes

$$(J \cdot \nabla)L_x + (J \cdot N)L_x, \text{ where } N = \nabla \times L.$$

We may interchange the order of the N and L_x operators in the second term by using the commutation relations of the form:

$$N_x L_y - L_y N_x = -N_z = -(N_y L_x - L_x N_y), \quad N_z L_x - L_x N_z = 0. \quad (3.12)$$

We find then that

$$(J \cdot N)L_x = L_x(J \cdot N) + (J \times N)_x$$

* It may be noticed that the operator in the square bracket in (3.4) automatically eliminates spherically symmetrical ($l = 0$) components of the operand. Since $L^{-2}(0)$ is not necessarily zero, conceivably the L^{-2} operator could restore terms of this type. However, the omission of $l = 0$ terms in what follows (e.g. in the function g) is a direct consequence of the earlier removal of the irrotational part of the vector potential.

from which it follows that

$$\nabla \times \{ \nabla \times [(J \times L) \times L] \} = N(J \cdot N);$$

the remaining terms cancel since

$$\nabla \times (J \times N) = - (J \cdot \nabla)N.$$

Therefore,
$$A_1 = -\frac{1}{k^2 c} N \int J \cdot N g dV'. \quad (3.13)$$

A still more symmetric form may be obtained if we use the fact that

$$NN_g = NN'_g^*$$

so that we may write finally

$$A_1 = -\frac{1}{k^2 c} N \int J \cdot N'_g dV' \quad (3.14)$$

The identification of (3.10) and (3.14) as the potentials for the magnetic and electric radiations respectively may be made by writing down the corresponding fields:

$$\begin{aligned} E_1 &= -\frac{i}{kc} N \int J \cdot N'_g dV', & H_1 &= -\frac{1}{c} L \int J \cdot N'_g dV', \\ E_2 &= -\frac{ik}{c} L \int J \cdot L'_g dV', & H_2 &= -\frac{1}{c} N \int J \cdot L'_g dV'. \end{aligned} \quad (3.15)$$

But now $r \cdot L = 0$, so that the radial components of H_1 and of E_2 are zero. If the radial magnetic field is zero, we describe the radiation as *electric multipole*; if the radial electric field is zero, it is *magnetic multipole*.

The integrals occurring in (3.10), (3.14), and (3.15) may be written in a more suggestive form. For

$$G_2 = -\frac{1}{c} \int J \cdot L'_g dV' = -\frac{2}{c} \int \text{div } M \cdot g dV' \quad (3.16)$$

where $M = \frac{1}{2} r' \times J$. We are also interested in

$$G_1 = \frac{i}{kc} \int J \cdot N'_g dV'.$$

Now
$$N' = \nabla' \times (r' \times \nabla') = r' \nabla'^2 - \nabla' (1 + r' \cdot \nabla') \quad (3.17)$$

Using this and (2.1) we get

$$G_1 = \int \rho (1 + r' \cdot \nabla') g dV' - \frac{i}{kc} \int (r' \cdot J) \nabla'^2 g dV'.$$

The second term bears to the first a ratio $\approx (ka)$, where a is the radius of the radiating system, and is consequently negligible in cases in which the dimensions of the radiating system are small compared with the wave length of the radiation.

*See Appendix.

From (3.15) the fields in the case of electric radiation are

$$\mathbf{E}_1 = N\mathbf{G}_1 \quad \mathbf{H}_1 = -ik\mathbf{L}\mathbf{G}_1, \quad (3.18)$$

and of the magnetic

$$\mathbf{E}_2 = +ik\mathbf{L}\mathbf{G}_2 \quad \mathbf{H}_2 = N\mathbf{G}_2. \quad (3.19)$$

The energy flux is given by

$$\mathbf{P} = \text{Re} \left\{ \frac{c}{8\pi} [(\mathbf{E}_1 + \mathbf{E}_2) \times (\mathbf{H}_1^* + \mathbf{H}_2^*)] \right\}$$

Where "Re" denotes "real part of", and * designates a complex conjugate.

But it may be verified directly that

$$\text{div}(\mathbf{E}_1 \times \mathbf{H}_2^*) = -\text{div}(\mathbf{E}_2 \times \mathbf{H}_1^*), \quad (3.20)$$

and so the contributions to the energy from the electric and magnetic radiations are independent of each other.

4. "Electric" Radiation

The energy flux of electric radiation over a sphere of radius r is

$$F_1 = \frac{c}{8\pi} \text{Re} \int r \mathbf{r} \cdot \mathbf{E}_1 \times \mathbf{H}_1^* d\Omega \quad (4.1)$$

where the integral is over solid angle. Now from (3.17)

$$\mathbf{r} \times \mathbf{N} = -\mathbf{L}\sigma = -\sigma\mathbf{L}, \text{ where } \sigma = 1 + r \frac{\partial}{\partial r}.$$

This enables us to write (4.1) as

$$F_1 = -\frac{ckr}{8\pi} \text{Re} \int \sigma \mathbf{L}\mathbf{G}_1 \cdot \mathbf{L}\mathbf{G}_1^* d\Omega, \quad (4.2)$$

$$= -\frac{ckr}{8\pi} \text{Im} \int \sigma \mathbf{G}_1 \cdot \mathbf{L}^2 \mathbf{G}_1^* d\Omega \quad (4.3)$$

where "Im" designates "imaginary part of". The integrand of (4.2) gives the radiation pattern, i.e. the radial energy flow as a function of position in general.

It is evident from this that the various electric multipole radiations are independent, since the spherical harmonics in terms of which we shall expand are eigenfunctions of \mathbf{L}^2 . Thus the respective contributions to the flux are independent.

Taking the imaginary part rather than the real part in (4.2) we could find (1) the difference between the mean electric and magnetic field energies. Since asymptotically G_1 and $G_2 \approx \frac{e^{ikr}}{r}$ this becomes zero in the "radiation region".

The *angular momentum* of the radiation is also easily found; the angular momentum density is

$$\begin{aligned} \mathbf{M}_1 &= \text{Re } \mathbf{r} \times \frac{1}{8\pi c} (\mathbf{E}_1 \times \mathbf{H}_1^*), \\ &= \frac{k}{8\pi c} \text{Im } \mathbf{L}^2 G_1 \cdot \mathbf{L} G_1^*. \end{aligned} \quad (4.4)$$

Thus the expansion in spherical harmonics will also be an expansion in the eigenvalues of angular momentum. Furthermore, the angular momentum density goes asymptotically as $\frac{1}{r^2}$, so that there is the same angular momentum in any shell $(r, r + dr)$. Therefore, the angular momentum is effectively all in the radiation field.*

5. The "Magnetic" Radiation

The results here parallel very closely those of the preceding section. The energy flux may be written

$$\begin{aligned} F_2 &= \frac{ckr}{8\pi} \text{Re } i \int \mathbf{L} G_2 \cdot \mathbf{L}(\sigma G_2^*) d\Omega, \\ &= \frac{ckr}{8\pi} \text{Im } \int G_2 \mathbf{L}^2(\sigma G_2^*) d\Omega \end{aligned} \quad (5.1)$$

Since σ operates only on the *radial* part of G_2 , an expansion in multipoles again results in a complete separation of the terms of various orders. Also the angular momentum density comes out to be

$$\mathbf{M}_2 = - \frac{k}{8\pi c} \text{Im } \mathbf{L} G_2 \cdot \mathbf{L}^2 G_2^*. \quad (5.2)$$

6. Multipole Expansion

Let us now introduce the "tensor spherical harmonics" (10)

$$Y_{i_1, i_2, \dots, i_l} = \frac{(-1)^l}{l!} r^{l+1} \frac{\partial^l}{\partial x_{i_1} \partial x_{i_2} \dots \partial x_{i_l}} \left(\frac{1}{r} \right), \quad (6.1)$$

and make use of the fact that

$$P_l(\cos \theta) = C_l Y_{(l)}(\Omega) Y_{(l)}(\Omega') \quad (6.2)$$

$$\text{where} \quad C_l = \frac{(l!)^2 2^l}{(2l)!} \quad (6.3)$$

The " (l) " indicates l dummy indices of summation and the angle θ is that between the directions Ω and Ω' . We may now use the expansion (3.6) for g ,

*However, there would be no angular momentum if the fields were both without radial components. The decrease in the radial component with r is compensated by the increase in the volume $4\pi r^2 dr$.

substituting from (6.2) for $P_l(\cos \theta)$, to obtain expansions in harmonics (corresponding to the various multipole orders*) for G_1 and G_2 ; the fields then follow from (3.18) and (3.19).

$$G_1 = -ik \sum_l \frac{2l+1}{l(l+1)} C_l h_l^{(1)}(kr) Y_{(l)} Q_{(l)} \quad (6.4)$$

$$\text{where} \quad Q_{(l)} = \int \left(\rho \sigma' + i \frac{k}{c} \mathbf{r}' \cdot \mathbf{J} \right) j_l(kr') Y_{(l)'} dV', \quad (6.5)$$

σ' being the operator $\left(1 + r' \frac{\partial}{\partial r'}\right)$.

The radiation intensity may therefore be written

$$F_1 = \sum_{l=1}^{\infty} F_1^{(l)} \quad (6.6)$$

where

$$F_1^{(l)} = \text{Im} \left[\frac{ck^2}{8\pi} k r h_l^{(1)*}(kr) \sigma h_l^{(1)}(kr) \left(\frac{(2l+1)C_l}{l(l+1)} \right)^2 Q_{(l)} Q_{(l')} \int Y_{(l)} Y_{(l')} d\Omega \right] \quad (6.7)$$

and the (l) , (l') refer to *different* sets of l dummy indices.

$$\text{Now} \quad \int Y_{(l)} Y_{(l')} d\Omega = \frac{4\pi}{l!(2l+1)C_l} \delta_{(l')}^{(l)} + \text{further terms} \quad (6.8)$$

where $\delta_{(l')}^{(l)}$ = the sum of the $l!$ different products of Kronecker δ 's involving one index from l and one from l' , and the "further terms" involve δ 's of at least two indices from the set (l) or (l') . These latter terms give no contribution since, by their definition, a $Q_{(l)}$ with a repeated index vanishes. Thus the flux at radius r is

$$F_1^{(l)} = \frac{ck^2}{2} \frac{(2l+1)C_l}{l(l+1)} |Q_{(l)}|^2 \text{Im} \{ k r h_l^{(1)*}(kr) \sigma h_l^{(1)}(kr) \}, \quad (6.9)$$

where $|Q_{(l)}|^2$ is the scalar product of $Q_{(l)}$ and its complex conjugate. For large r , the radial factor approaches unity.

Similar calculations with the "magnetic" terms lead to

$$G_2 = \frac{2ik}{c} \sum_l \frac{2l+1}{l(l+1)} C_l h_l^{(1)}(kr) Y_{(l)} \mu_{(l)}, \quad (6.10)$$

where

$$\mu_{(l)} = \int \text{div} \mathbf{M} \cdot j_l(kr') Y_{(l)'} dV' \quad (6.11)$$

It follows that the flux of the magnetic 2^l -pole radiation has intensity

$$F_2^{(l)} = \frac{2k^3}{c} \frac{(2l+1)C_l}{l(l+1)} |\mu_{(l)}|^2 \text{Im} \{ -k r h_l^{(1)}(kr) \sigma h_l^{(1)*}(kr) \} \quad (6.12)$$

*It should be noted that, when the expansion is carried out at this point, $l = 1$ corresponds to dipole radiation (electric or magnetic), $l = 2$ to quadrupole, etc.

where the last term again becomes unity at large distances. It must be emphasized that the formulae (6.9) and (6.12) are valid without any approximations concerning the relative magnitude of the dimensions of the radiating system and the wave length.

If we are interested in the components of angular momentum, the tensor approach is probably no longer the natural one. Rather, it seems desirable, if one is to calculate the z -component of angular momentum, to use spherical harmonics in terms of polar coordinates with the z -axis as polar axis; i.e. to expand in terms of the eigenfunctions of L^2 and L_z .

It is also of some interest to calculate the *energy density*. For the "electric" type of radiation this is

$$E_{(e)} = \frac{1}{16\pi} [NG_1 \cdot NG_1^* + k^2 LG_1 \cdot LG_1^*]$$

Integrating this over volume, integrating by parts and using the fact that $N^2 = L^2 \nabla^2$

$$E_{(e)} = - \frac{k^2}{8\pi} \int G_1 L^2 G_1^* dV, \quad (6.13)$$

and for the "magnetic" radiation

$$E_{(m)} = - \frac{k^2}{8\pi} \int G_2 L^2 G_2^* dV. \quad (6.14)$$

If now we expand G_1, G_2 in terms of "conventional" spherical harmonics

$$Y_{lm} = \sqrt{\frac{(2l+1)}{4\pi} \frac{(l-|m|)!}{(l+|m|)!}} P_l^{|m|}(\cos \theta) e^{im\phi}, \quad (6.15)$$

designating the coefficients $G_{1,2}^{(lm)}$ for brevity, it is evident that the (lm) contributions to the z -component of angular momentum are

$$M_z^{(lm)} = \frac{k}{8\pi c} l(l+1) m \int |G_{1,2}^{(lm)}|^2 dV \quad (6.16)$$

while the corresponding contributions to the energy are

$$E^{(lm)} = \frac{k^2}{8\pi} l(l+1) \int |G_{1,2}^{(lm)}|^2 dV. \quad (6.17)$$

Thus we have the general result

$$M_z^{(lm)} = m \frac{E^{(lm)}}{\omega}. \quad (6.18)$$

For completeness, we list the expansion coefficients

$$G_1^{(lm)} = 4\pi i k \frac{1}{l(l+1)} Q_{lm} h_l^{(1)}(kr) Y_{lm}(\theta, \varphi), \quad (6.19)$$

$$\text{where} \quad Q_{lm} = \int \left[\rho \sigma' + i \frac{\hbar}{c} \mathbf{r}' \cdot \mathbf{J} \right] j_l(kr') Y_{lm}(\theta', \varphi') dV' \quad (6.20)$$

$$\text{and} \quad G_2^{(lm)} = -\frac{2}{c} 4\pi i k \frac{1}{l(l+1)} \mu_{lm} h_l^{(1)}(kr) Y_{lm}(\theta, \varphi), \quad (6.21)$$

$$\text{where} \quad \mu_{lm} = \int \operatorname{div} \mathbf{M} \cdot \mathbf{j}_l(kr') Y_{lm}(\theta', \varphi') dV'. \quad (6.22)$$

Both the total angular momentum and total energy will of course be finite only if the system has been radiating only for a finite period of time.

Appendix

We may write (cf. Equation 3.5)

$$\mathbf{L}g = \mathbf{L}'^{-2} \mathbf{L} \frac{e^{ikR}}{R} \quad (A.1)$$

where \mathbf{L}'^{-2} is the inverse operator to \mathbf{L}'^2 . Then

$$\begin{aligned} \mathbf{N}g &= \mathbf{L}'^{-2} (\nabla \times \mathbf{L}) \frac{e^{ikR}}{R}, \\ &= -\mathbf{L}'^{-2} (\mathbf{L}' \times \nabla') \frac{e^{ikR}}{R}, \\ &= -\mathbf{L}'^{-2} (\mathbf{L}' \times \nabla') \mathbf{L}'^2 g + \left\langle \frac{e^{ikR}}{R} \right\rangle_{\mathbf{A}\mathbf{v}}. \end{aligned} \quad (A.2)$$

It is now necessary only to know the commutation rule between ∇' and \mathbf{L}'^2 :

$$\nabla' \mathbf{L}'^2 - \mathbf{L}'^2 \nabla' = 2\mathbf{N}' + 2\nabla'$$

and that between ∇ and \mathbf{L} :

$$\begin{aligned} \nabla_x L_y - L_y \nabla_x &= -\nabla_z = -(\nabla_y L_x - L_x \nabla_y), \\ \nabla_x L_x - L_x \nabla_x &= 0, \text{ etc.} \end{aligned}$$

to verify that $(\mathbf{L}' \times \nabla') \mathbf{L}'^2 = -\mathbf{L}'^2 (\nabla' \times \mathbf{L}')$

Multiplying by \mathbf{L}'^{-2} on the left, we get

$$\mathbf{L}'^{-2} (\mathbf{L}' \times \nabla') \mathbf{L}'^2 = -\mathbf{N}'.$$

Putting this in (A.2) and operating with another \mathbf{N} gives the required result:

$$\mathbf{N}\mathbf{N}g = \mathbf{N}\mathbf{N}'g. \quad (A.3)$$

References

1. ABRAHAM, M. AND BECKER, R. Classical theory of electricity and magnetism. Blackie and Sons, Ltd., London and Glasgow. 1932.
2. BERESTETZKY, V. J. Phys. U.S.S.R. 11: 85. 1947.
3. FIERZ, M. Helv. Phys. Acta, 22: 489. 1949.
4. FRANZ, W. Z. Physik, 127: 363. 1950.
5. HEITLER, W. Proc. Cambridge Phil. Soc. 32: 112. 1936.
6. HUMBLET, J. Physica, 10: 585. 1943.
7. HUMBLET, J. Physica, 11: 100. 1944.
8. KRAMERS, H. A. Physica, 10: 261. 1943.
9. LLOYD, S. P. Phys. Rev. 81: 161. 1951.
10. WALLACE, P. R. Can. J. Research, A, 26: 99. 1948.

VARIATION OF METEOR ECHO RATES WITH RADAR SYSTEM PARAMETERS¹

By D. W. R. MCKINLEY

Abstract

Observations made with crossed-polarization radar system systems do not support the suggestion that the ionized meteor trail may act as a strong filter-polarizer of the incident radio wave. Experiments have been carried out to determine the variation of normal meteor echo rates with transmitter power, antenna gain, and radio wave length, and all confirm Lovell's scattering formula, provided that account is taken of the effective broadening of the scattering pattern of the meteor trail with increasing wave length. The limiting sensitivity of the 9.22 m. 200 kw. radar is determined to be about 9th magnitude. During a strong visual shower the observed increase in visual rates and low-power radar rates, compared to high-power radar rates, is explained by assuming that the magnitude distribution of the shower meteors differs from the normal nonshower distribution.

1. Introduction

From earlier work by Blackett and Lovell (2), Lovell and Clegg (8) have developed a theoretical relation between the electron density in a meteor trail and the radar system parameters, assuming the echo signal to be produced by scattering from a long line of free electrons. When the numerical values for the physical constants are inserted this formula becomes,

$$\epsilon = 3.3 \times 10^{-28} \frac{a^2 \lambda^3}{R^3} PG^2 \text{ w.}, \quad (1a)$$

$$\text{or, } a = 5.5 \times 10^{13} \frac{1}{G} \left(\frac{\epsilon R^3}{P \lambda^3} \right)^{1/2} \text{ electrons per cm.} \quad (1b)$$

where ϵ = received echo power in watts

a = number of electrons produced per cm. of path

G = antenna gain

R = distance to meteor trail in cm.

P = peak transmitted power in watts

λ = wave length of transmitter in cm.

Lovell (8) demonstrated that the predicted variation of ϵ with λ was in agreement with the observed variation over the range $\lambda = 1.4$ to 8.3 m., for individual echoes of the type that could generally be associated with visual meteors. In an earlier paper (7) he showed that the number of observed echoes falling in a given interval about a selected value of a was approximately inversely proportional to a . From theoretical considerations, Herlofson (5) indicated that a 1st magnitude visual meteor moving at 40 km. per sec. may create 10^{12} electrons per cm. of path and a 6th magnitude meteor about 10^{10} electrons.

¹ Manuscript received March 7, 1951.

Contribution from the Radio and Electrical Engineering Division, National Research Laboratories, Ottawa, Canada. Issued as N.R.C. No. 2502.

However, Lovell (9) has suggested that the formula may not be applicable to the brighter meteors detected on longer wave lengths. The question first arose in comparing the rates of meteor echoes on 8.3 m. (and on 9.2 m. (11)) to the rates of echoes on 4.2 m. and shorter wave lengths. There appeared to be a sudden increase in the hourly counts for wave lengths greater than about 6 m., and Lovell (10) suggested that other scattering processes may also occur on the longer wave lengths. He interpreted the phenomenological theory of McKinley and Millman (11), based on their observations on 9.2 m., as implying that a multiplicity of echoes can arise from a single meteor. Herlofson (4) recently proposed a theory of plasma oscillations in the column of ionized gas formed in the wake of the meteor. One consequence of this theory was that under certain conditions obtaining on the longer wave lengths the echo strengths could be many magnitudes greater than would be expected from the formula, Equation 1.

Alternatively, one might regard the ionized cloud left by the meteor as a miniature ionosphere. If the cloud is assumed to have an effective echoing area, $A \text{ cm}^2$, the received signal power may be found from the conventional radar equation (14),

$$\epsilon = \frac{PG^2A\lambda^2}{(4\pi)^3R^4} \quad (2)$$

The volume density of electrons must exceed the critical density for the wave length employed, otherwise no echo will be received regardless of the size of the echoing area. The critical density is inversely proportional to the square of the wave length. As the wave length is increased fainter meteors should be detected, both for this reason and also because of the λ^2 factor in Equation (2). In effect, the minimum detectable echo power should depend on λ^4 for this total reflection case, compared with λ^3 for the case involving scattering from a long line source of free electrons.

In the course of a general program of meteor observing, which has been in progress at Ottawa since the summer of 1947 under the combined auspices of the Dominion Observatory and the National Research Council, some experiments have been conducted on 9.2, 5.4, and 2.8 m., with various transmitter powers and antenna systems. These have led to the broad conclusions that Lovell's formula applies reasonably well to the majority of meteors detected on wave lengths as long as 9.2 m. However two new factors are brought forward which have a bearing on the variation of meteor echo rates. One is a further decrease in the number of detectable echoes as the wave length is decreased; this effect is a consequence of the geometry of the diffraction pattern associated with the scattered echo from the meteor trail, and has not been included in Equations 1a, 1b. The other is a distinct difference in the size distribution of meteors belonging to a strong shower as compared with the distribution of the normal background meteors. The following sections, in order, describe experiments designed to test Herlofson's plasma theory, the

variation of echo rates with antenna gain, the variation of rates with transmitter power, the variation of rates with wave length, and the general conclusions.

2. Polarization Effects

Herlofson's theory (4) of plasma oscillations suggests that, if the volume density of electrons in the meteor trail is sufficient to cause the dielectric constant of the medium to become negative, an anomalous effect should be observed which is dependent on the polarization of the incident electromagnetic wave relative to the axis of the meteor trail. If the electric vector is parallel to the trail the usual scattering processes apply, defined by Equation (1). If the electric vector is normal to the ionized column, plasma or resonance oscillations are presumed to occur, with the result that the scattering should become *several powers of 10* stronger than for the parallel case. Herlofson adds that, owing to this resonance phenomenon, an ionized column would act as a combined filter and polarizer for the scattered radiation. It would return with great strength radiation with such low frequency that the dielectric constant is negative in the center, but would respond feebly to radiation of higher frequencies. On the favorable frequencies only the wave component with electric vector normal to the cylinder would be scattered with appreciable intensity.

If the incident electric field vector is at an angle β to the meteor column it may be resolved into components $\cos \beta$ and $\sin \beta$ which are respectively parallel and normal to the cylinder, and we shall assume from the above discussion that only the normal component will contribute significantly to the received echo signal. Our experience (11) has shown that the great majority of meteor echoes observed on 9.2 m. occur at or near the point where the meteor path is normal to the line of sight: the brighter meteors can certainly be detected over much wider aspects but their number is statistically insignificant. For the majority of echoes, therefore, the angle β will lie in the plane of the incident wave front, which simplifies the geometry of the situation. If observational periods are selected when no major meteor showers are active it is reasonable to assume that the sporadic meteor paths will be in random directions; that is, there will be a fairly random distribution of the angle β .

Case I.—Consider two similar but independent radar systems, one operating with its transmitting and receiving dipoles oriented North-South, and the other East-West. Each antenna is a half-wave horizontal dipole mounted a quarter-wave above a ground screen reflector. The wave lengths differ slightly to avoid mutual interference. The electric vector of the N/S transmitter, expressed in volts per cm. at the target distance, will produce after reflection from the meteor column a scattered field strength proportional to $\sin \beta$, the vector being normal to the column. This vector will have a component proportional to $\sin \beta$ in the direction of the receiving dipole of the N/S system so that the resultant voltage produced at the input of the receiver will be proportional to $\sin^2 \beta$. Similarly the receiving dipole of the E/W system will be

excited by a field proportional to $\cos^2 \beta$. If the receiver outputs vary linearly with field strength the output signals will be respectively proportional to $\sin^2 \beta$ and $\cos^2 \beta$, and the ratio of echo voltages on the two systems will be proportional to $\tan^2 \beta$.

Case II.—Consider one radar transmitter with its dipole oriented N/S, and two receivers both tuned to the transmitter frequency, one having its dipole N/S, and the other E/W. The N/S receiver will produce echoes proportional to $\sin^2 \beta$ as before, but the E/W receiver output will now be proportional to $\sin \beta \cos \beta$ and the echo voltage ratios will be proportional to $\tan \beta$.

One of the basic requirements of the two cases described above is that the electric vectors of the N/S and the E/W systems shall be at right angles to each other at the meteor target, and this condition holds accurately only in the

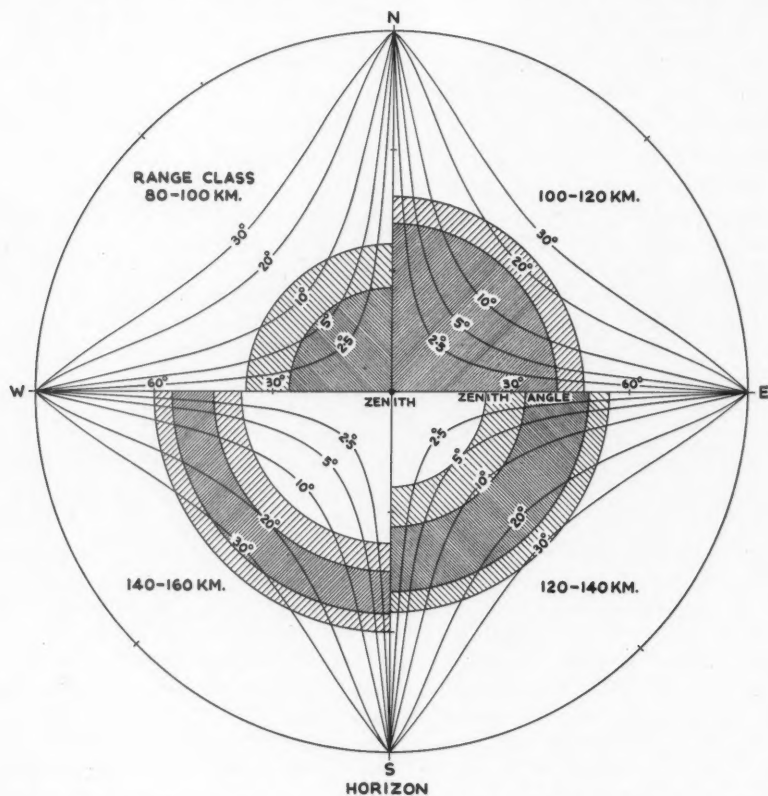


FIG. 1. Zenithal projection of the sky. The shaded areas show the meteor zones corresponding to the selected range classes. The curves, marked 2.5° to 30° , indicate the departure from orthogonality of the N/S and E/W systems.

zenith and along the two great circles passing through the zenith and projected axes of the dipoles.

Fig. 1 is a zenithal equidistant projection of the sky on which have been drawn curves to indicate the departure from orthogonality of the electric vectors of the two systems, the selected values being 2.5° , 5° , 10° , 20° , 30° . On assumption that the part of the M-region lying between 90 km. and 100 km. height above sea level is more productive of meteor echoes (11, 13), heavily shaded areas have been drawn to show the zones of the sky from which most of the echoes in a given radar range class may be expected to occur. The lightly shaded areas denote the height regions, 80-90 km., and 100-110 km., from which fewer echoes are expected. Because the range-time records of the radar systems carry 20 km. calibration lines it is convenient to consider range classes in steps of 20 km. The sky coverage for the range class 80-100 km. is shown in the upper left quadrant only of Fig. 1, that for the range class 100-120 km. in the upper right quadrant, and so on. It will immediately be apparent that the range class 80-100 km. is the best one to consider for our purposes as the departure from orthogonality for most of the echoes should be less than 2.5° and for very few as great as 10° . The range class 100-120 km. embraces a substantial region where the departure is in excess of 5° but the statistics for this class should still be of some value. For the range classes 120-140 km. and greater the simple analysis breaks down, and, in fact, for the long range classes at low angles the polarizations of the two systems are not far from being parallel for a great percentage of the echoes.

A further reason, and a very practical one, for rejecting data from the long range classes was the marked difference in the antenna coverage of the N/S and the E/W systems at low elevation angles. There is a significant difference in the theoretical radiation patterns of a dipole backed by a reflecting screen, depending on whether the measurement is made in the plane of the dipole or normal to it, though this difference is of importance only at low angles. Furthermore, under the actual experimental conditions, several towers, buildings, poles, and guy-wires in the neighborhood of the transmitting and receiving dipoles caused additional distortions in the antenna patterns at low angles. The distortions were not the same for the two systems because of the different effects of the polarizations of the various reflecting objects in the immediate vicinity, particularly as the antennae were separated by a few yards. It is unlikely, though, that any serious difference in coverage existed at high angles.

For a study of Case I two transmitters were used which were similar but not quite identical. The E/W transmitter used two tubes in the oscillator with a power output of about 200 kw. peak, whereas the N/S oscillator used four tubes and the measured relative power output was 40% greater. Both transmitters were keyed in parallel from the same modulator with the same pulse width, pulse voltage, and recurrence frequency. To avoid mutual interference the N/S system operated on 9.22 m. and the E/W system on 9.52 m. With this separation each receiver had a rejection ratio of more than 90 db. for un-

desired to desired echo signals. In comparing echo strength on the two systems a small correction factor should be applied. Employing Lovell's formula (Equation (1a)), the difference in the systems amounted to a power ratio of 1.25 or 1 db. in favor of the N/S system, which was negligible in comparison with the probable error of 6 db. involved in estimating most of the ratios of echo amplitudes on the two systems. Equation (1a) may not be the correct relation to use, a priori, but the differentials in power and wave length are small and tend to cancel, hence it is unlikely that any other reasonable formula (Equation (2) for example) would show any significant difference.

Two identical receivers were used, and the outputs displayed on two cathode ray tubes mounted side by side and photographed simultaneously by a camera in which the film was drawn in a vertical plane continuously past the open lens at 2 in. per minute. The same horizontal range sweep and calibration marks were applied to each C.R. tube and the receiver outputs were fed to the intensity grids of the C.R. tubes. During the runs the N/S and E/W antennas were interchanged frequently between the receiver-display systems to eliminate any discrepancies in performance, though care was taken to equalize them initially as far as possible. The video output of each receiver passed through an adjustable clipper circuit which could be set to limit the amplitude of strong signals at any given level, before application to the C.R. tube.

In these experiments, the background noise level output of the receivers was set at the optimum value for operation of the displays, and the clipping level was set at twice the mean noise level, measured in volts on an A-scan monitor C.R. tube. The receiver band widths were greater than the minimum required for the transmitter pulse width (20 μ sec.) so that weak echoes were easily detectable by virtue of the shape of the echo spot compared to the finer-grained noise spots. On the average, echoes of half the mean noise voltage was clearly resolvable on the film. Thus, the working range of the receiver between the minimum detectable signal and a signal just reaching the clipping level was 4:1 in voltage, or 12 db. Signals above the clipping level produced a black shadow, increasing in range on the display in proportion to the signal strength, due to the video overshoot effect in the clipping circuits. The variation of this shadow length with input signal voltage could be calibrated fairly accurately up to the point where the receiver began to saturate prior to the clipping circuits. We arbitrarily defined 0 db. as the absolute power level of the weakest detectable signal; therefore 12 db. was the absolute power of an echo just reaching the clipping level. It was found that 24 db. represented the strongest signal that could be measured before the receiver characteristic became nonlinear. Signals below 0 db. were not detected, of course, and signals above 24 db. could not be accurately measured, but there were very few above the 24 db. level.

For statistical purposes it was therefore convenient to divide the working range of a receiver into two parts, the first extending from 0 to 12 db. and including all signals from the weakest perceptible signal to one just touching

the clipping level, and the second including all echoes over the 12 db. absolute power level. These signal strength groups may be called *absolute* classes, based on the arbitrary zero defined above, to distinguish them from the *relative* classes which we shall use to classify the observational data. That is, we shall endeavor to list all the echoes as falling either into a relative class *under 12 db.*, wherein the *difference* in the echo strengths on the two displays is less than 12 db., or into a second relative class *over 12 db.*, which includes all the remainder.

In the reduction of the records the echoes were considered in pairs, using range to the nearest km., time to 1/10th of a second, and echo shape to correlate the pairs. If an echo pair appeared on both the N/S and E/W records, each with a signal strength between 0 and 12 db., the pair was assigned to the *under* class immediately. If one echo was below the clipping level and the other above it a more careful estimate of the relative signal strengths was required before classifying the pair as either *under* or *over*. If both echoes were above 12 db. the pair was classed as *under*: a few of these should have been labeled *over*, where one or both echoes of the pair reached saturation, but the number of this kind was statistically insignificant.

When one member of a pair was missing the pair was classed *over*. In this case, if the observed echo lay between 0 and 12 db., the missing echo might have been just below detectability but still within 12 db. of the observed echo. The probable number of such pairs which should have been marked *under* was by no means negligible, as most of the observed echoes lay close to the 0 db. absolute level, but it was difficult to determine the true proportion without making assumptions in advance regarding the distribution of echoes with signal strength. Without applying such a correction it should be noted that this method of analysis will put too many echoes in the *over* class. Table I summarizes the echo classification techniques.

In Case I the theoretical voltage ratio of signals on the two systems is expressed as $\tan^2\beta$. For up to 12 db. difference in echo power (4:1 in field strength or echo voltage), β can lie between the value $\tan^{-1}2$ and $\tan^{-1}1/2$. Thus, the fraction of echoes to be expected in the *under* class is the ratio

TABLE I
ECHO CLASSIFICATION

Echo type	Absolute class of observed echo		Relative class of echo pair (<i>over</i> or <i>under</i> a 12 db. difference in signal strength)
	Either receiver	The other receiver	
1	0-12 db.	No echo	<i>Over</i>
2	0-12 db.	0-12 db.	<i>Under</i>
3	Over 12 db.	No echo	<i>Over</i>
4	Over 12 db.	0-12 db.	<i>Under</i> or <i>over</i> (see text)
5	Over 12 db.	Over 12 db.	<i>Under</i>

$(\tan^{-1}2 - \tan^{-1}1/2)$ to $\pi/2$. Expressed in per cent, the simple filter-polarizer theory predicts that 41% should be listed as *under* and 59% as *over*. In Case II, where the voltage ratio is proportional to $\tan \beta$, the corresponding angular ratio is $(\tan^{-1}4 - \tan^{-1}1/4)$ to $\pi/2$; whence it follows that 69% should be *under* and 31% *over*. If we now go to the other extreme and postulate that no polarization effects whatsoever are introduced by the mechanics of reflection from the meteor trail, i.e., that the plane of polarization of the reflected wave remains in the original plane, then it follows immediately that for Case I 100% of the echoes should be listed as *under* and none as *over*, while for Case II none should be *under* and 100% should be *over*. As another alternative if we assume that every echo is circularly polarized then in both Case I and Case II 100% of the echoes should be *under*. The "no polarization" and the "circular polarization" hypotheses are of academic interest but may be useful in suggesting limits.

The two independent transmitter-receiver systems (Case I) were operated from 1100 E.S.T. to 1500 E.S.T. on Nov. 26, 1948, and from 1700 E.S.T. Dec. 2, to 1200 E.S.T. Dec. 3, 1948. A part of the films was discarded because interference from local and distant sources made it difficult at times to estimate relative signal strengths accurately. About seventeen hours of satisfactory records were analyzed. The echo coincidences were noted in pairs, according to the procedure of Table I, and the numbers of *under* and *over* echo pairs are shown in Table II, Row 1, with the percentages given in Row 2. In the *over* classes the numbers of Type I pairs are indicated in brackets. Row 3 contains the predicted percentages based on the filter-polarizer theory, Row 4 the percentages assuming no polarization effects, and Row 5 the percentages for complete circular polarization.

The observed percentages of echoes in the *over* class should be reduced appreciably. Most of the echoes with both members of the pair showing are in the *under* class, hence it is reasonable to expect that a large fraction of the Type I echoes should also be in the *under* class. However, even without attempting to adjust the figures in this respect, it is seen that the wide divergence between the

TABLE II
ECHO DISTRIBUTIONS FOR CASE I

	80-100 km. range class		100-120 km. range class	
	<i>Under</i>	<i>Over</i>	<i>Under</i>	<i>Over</i>
1. Observed numbers	402	53 (44)	1837	322 (208)
2. Uncorrected percentages	88%	12%	85%	15%
3. Filter-polarizer theory	41%	59%	41%	59%
4. No polarization	100%	0%	100%	0%
5. Circular polarization	100%	0%	100%	0%

observed percentages of *over* and *under* classes and the percentages predicted by the filter-polarizer theory indicates that the meteor trail does not act as a combined filter and polarizer of the incident radio wave, at least, not to the extent that the component normal to the trail is several orders of magnitude stronger than the component parallel to the trail. It would be inadvisable to conclude that these observations refute the possibility of plasma oscillations occurring in the meteoric cloud; they merely imply that the deductions made from the plasma hypothesis are not correct for the majority of meteor echoes. On the other hand the data presented so far cannot be said to favor either of the extreme hypotheses, since both predict 100% *under* in Case I. More information is required, some of which may be available from a study of Case II.

The N/S transmitter only was operated with both the N/S and E/W receivers tuned to the transmitter wave length of 9.22 m. (Case II) for a total of 13 hr. on the nights of Mar. 25 and Mar. 30, 1949. From these records six hours were selected which were satisfactorily free of interference. Table III summarizes the observations, together with the various theoretical predictions for this case.

TABLE III
ECHO DISTRIBUTIONS FOR CASE II

	80-100 km. range class		100-120 km. range class	
	<i>Under</i>	<i>Over</i>	<i>Under</i>	<i>Over</i>
1. Observed numbers	41	62 (26) *	228	425 (227)
2. Uncorrected percentages	40%	60%	35%	65%
3. Filter-polarizer theory	69%	31%	69%	31%
4. No polarization	0%	100%	0%	100%
5. Circular polarization	100%	0%	100%	0%

Here, again the Type I pairs, shown in brackets, contain some that should have been listed in the *under* class, though the proportion would be less than in the preceding case because the proportion of known *under* class echoes is less. If a correction were applied it would tend to narrow the gap between the observed percentages and the percentages predicted by the filter-polarizer theory, rather than to widen it as in Case I. One point that the tabulated data does not emphasize is that the majority of the echo pairs in both the *under* and the *over* classes were stronger on the N/S receiver. In fact, there were only three echoes in six hours, in both the 80-100 km. and the 100-120 km. range classes, that appeared on the E/W receiver and not on the N/S receiver. The results in Case II should therefore be regarded as inconclusive as far as the filter-polarizer theory is concerned.

Considering the data for Case I and Case II together it appears evident that the meteor trail does not act as a filter-polarizer with an efficiency of

several orders of ten as suggested by Herlofson, though the observations do not exclude the possibility that the effect might be present but of much smaller magnitude, e.g., a factor of five, or that it might be applicable only to a few meteors under special circumstances. However, comparing the observations with the "no polarization" percentages in Table II and Table III, it seems probable that the returning wave is largely polarized in the incident plane (certainly it is not circularly polarized), with an appreciable content of random polarization.

3. Variation of Rates with Antenna Gain

As a general rule the transmitter and the receiver of the 9.2 m. radar system at Ottawa have each been coupled to a separate single half-wave folded dipole, mounted a quarter-wave above a reflecting ground screen (rather than using the same antenna to transmit and receive). For some observations on Mar. 9 and 10, 1948, three parasitic director elements were added to each of the antennae to increase the directivity vertically upwards. It was not practical to measure directly the relative power gains of the two 9.2 m. antenna systems, hence scaled-down models of the antennae were built for a wave length of 1.5 m. The antenna with three directors was determined to have a power gain of 2.2 over the antenna without the directors, measured at a point several wave lengths distant and vertically above the antenna.

The original intention of the experiments with the directive antenna system was to look for meteor echoes that might occur well above the M-region (11), possibly in the upper parts of the E-region or in the F-regions. No conclusions could be reached because any evidence of echoes that might have occurred at heights of 180–250 km. overhead was masked by the number of echoes detected at low angles in the side lobes of the antenna pattern at slant ranges of 180–250 km. More recent experience has indicated that the existence of an appreciable number of meteor echoes at the higher altitudes is very unlikely (13). However, the experimental data obtained with the high gain antenna may be of some interest here in correlating echo rates with antenna gain.

Mean hourly echo rates for each of the 20-km. range classes from 80 to 320 km. were obtained from a representative six-hour period of observing with the high-gain antenna, namely 0300–0900 E.S.T. Mar. 10, 1948 (see Fig. 2, dotted curve). The minimum at 150 km. corresponds to the first null in the antenna pattern and the smaller maxima to the side lobes. For comparison purposes similar data obtained with the low-gain antenna are shown plotted as the solid line in Fig. 2, averaging the periods 0300–0900 E.S.T. Feb. 6, 1948, and 0300–0900 E.S.T. Apr. 22, 1948. Observations with the low-gain antenna should preferably have been carried on simultaneously with the high-gain experiments. While the transmitter and receiver parameters were substantially the same for all three periods there was no immediate assurance that the distribution and frequency of meteors would be similar on the three dates. However, evidence accumulated during 1949 and 1950 (12) does suggest that the normal

background meteor rates for a given hour of the day (excluding the effects of the strong meteor showers) change slowly from day to day. It would therefore be reasonable to expect that the mean of the rates for the periods available in February and April should be representative of the rates for March with the same antenna system.

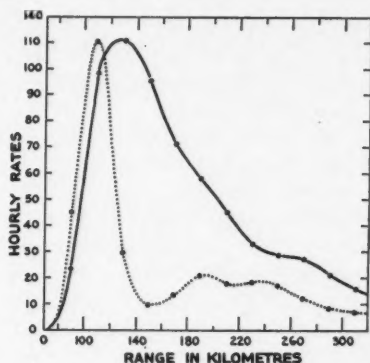


FIG. 2. Range distribution curves.
Solid line—low-gain antenna.
Dotted line—high-gain antenna.

The echo rates for the low-gain and high-gain antennae should be compared only for the range class 80–100 km. As will be seen from Fig. 2 dashed curve the gain of the multiple-element antenna fell off rapidly to a minimum for echoes in the range class 140–160 km. and was appreciably less than the maximum for the range classes 100–120 km. and 120–140 km. The observed mean hourly rates for the 80–100 km. class were 45 echoes per hour for the high-gain antenna and 24 echoes per hour for the low-gain antenna, or a ratio of 1.9 to 1. This figure is in fair agreement with the measured gain ratio of 2.2 for the two antennae, in view of the possible sources of error.

The mean hourly rates, summed over all range classes, for the high-gain and low-gain antennae respectively were 307 and 628 echoes per hour. Although the high-gain antenna system can detect smaller meteors the low-gain antenna covers a greater volume of the M-region, and hence detects a larger total of meteors despite the decrease in echo strength with increasing range. If the high-gain antenna were tilted to 45° , say, the echo rate should increase, both because the antenna pattern would intercept a larger volume and because meteors coming from radiants as high as 45° above the horizon could then be detected, whereas the radiants of detectable meteors in the overhead zone must lie close to the horizon. In comparing the background echo rates obtained with different equipments not only do the antenna patterns have to be specified but also the elevation of the maximum of the antenna beam. Furthermore, if a strong shower is present the antenna beam position must be completely specified; unless the beam lies in a plane normal to the paths of the shower meteors

very few of the shower meteors will be detected (6, 7). In all further work presented in this paper the single dipole antenna was used, and the conclusions reached will apply specifically to such an antenna only.

4. Variation of Rates with Transmitter Power

A series of tests was performed to ascertain the dependence of echo rates on transmitter peak power. For this purpose the d-c. voltage of the power supply to the modulator was varied in discrete steps, and the relative radiofrequency power outputs were measured by means of a r.f. current meter at the end of a quarter-wave stub line connected across the transmission line to the antenna. The relative powers could thus be measured with an accuracy of 5% or better. The absolute powers were known to an accuracy of about 20% only, because the computation of absolute power required a knowledge of the pulse width and shape, the transmission line surge impedance, and the pulse repetition frequency. No other parameters of the system were changed during the experiments, except that the receiver was kept tuned to the transmitter frequency (when the power was varied by as much as 16 to 1 the transmitter shifted frequency by 0.1 Mc.). During a series of observations the transmitter was operated on high and low power alternatively for half-hour periods and the echo counts averaged over several hours. Two experiments were carried out, one with a high-low power ratio of 6.25 to 1 (current ratio 2.5 to 1), see Table IV, and the other with a power ratio of 16 to 1 (current ratio 4 to 1), see Table V. The echo counts were made for each 20-km. range class from 80 km. to 260 km. (80 denotes the class 80-100 km., etc.) and the high-low echo ratios are given for each range class as well as for the totals.

It will be seen that the high-low ratios of total echo counts in Tables IV and V are in reasonable agreement with the transmitter r.f. current ratios, or, that the echo rates are proportional to the square root of the transmitter power. It will be noticed that the values of the ratios for the individual 20 km. class intervals do not show any definite trend from 80 km. to 260 km. If we tentatively adopt Lovell's formula (Equation 1*b*) and assume from the calculated antenna pattern that the antenna gain at the elevation angle of the 240-260 km. range class is about one-tenth the gain at 95 km. overhead, we obtain a

TABLE IV
HIGH-LOW POWER TEST
1900-0415 E.S.T. OCT. 30-31, 1950. CURRENT RATIO 2.5 TO 1
HIGH POWER 225,000 W. LOW POWER 36,000 W.

Range class, km.	80	100	120	140	160	180	200	220	240	Total
Hourly rates, high power	27.6	142.8	156.4	121.2	80.0	59.2	47.2	25.2	17.2	676.8
Hourly rates, low power	14.4	65.6	57.6	40.0	31.6	23.6	18.4	10.8	6.0	268.0
High-low ratio	1.92	2.18	2.72	3.03	2.53	2.51	2.57	2.33	2.87	2.53

TABLE V
HIGH-LOW POWER TEST
0000-0400 E.S.T. NOV. 29, 1950. CURRENT RATIO 4.0 TO 1
HIGH POWER 155,000 w. LOW POWER 9700 w.

Range class, km.	80	100	120	140	160	180	200	220	240	Total
Hourly rates, high power	22.5	96.5	130.0	140.0	115.5	85.5	53.5	41.0	33.5	718.0
Hourly rates, low power	3.5	31.5	32.0	36.0	28.0	21.5	18.5	11.5	7.0	189.5
High-low ratio	6.4	3.1	4.1	3.9	4.1	4.0	2.9	3.6	4.8	3.79

theoretical ratio of approximately 40 to 1 for the minimum detectable electron densities at the two ranges. This would be equivalent to a range of four visual magnitudes, if we accept Herlofson's proposed relationship. The significance of the comparatively constant high-low ratios in Tables IV and V as the range is increased is that the law of numbers versus size of the meteors in the long range classes must be the same as that for the meteors in the short range classes which are several magnitudes fainter on the average.

In Fig. 3 are plotted total hourly echo rates against the square root of the transmitter power in watts. The circled points were obtained from Table IV and the crosses from Table V. For each of the cases a straight line may be drawn through the points and the origin. The lines have not the same slope because the data from Table IV include observations made before midnight when the echo rates are lower. If the hours after midnight only are used from the October period the high and low hourly rates are found to be 856 and 334 respectively, and these selected values are plotted as crossed circles which lie fairly close to the line drawn through the November data. This implies that the normal echo rate for a given hour of the day changes but slowly from day to day, in the absence of a strong meteor shower; a conclusion which has been borne out by several hundred hours of observation extending over 15 months, made with a continuous-wave radio technique (12). An error in the parameters involved in computing the absolute power units used in Fig. 3 will cause the abscissa scale to expand or contract without changing the relative ratios of the plotted points.

The large transmitter could not be operated successfully for more than a 16 to 1 variation in power. Accordingly, a small radar transmitter was built to determine the ratio of echo rates for a large difference in power, particularly to compare the echo rate ratio obtained during a shower period with that obtained in a nonshower, or normal period. During the observing periods the low-power radar was maintained at 100 w. output on 9.52 m., and the high-power radar on 9.22 m. was held as closely as possible to 200 kw. output. Separate and identical receiving antennae, receivers, and displays were employed in the arrangement used for the crossed-polarization experiments, except that all the dipoles were oriented N/S. The 90 db. rejection ratio of the two receivers was thought to be sufficient to ensure that no echoes from the high-

power transmitter would appear on the low-power display, but as an additional precaution the low-power transmitter was triggered 133 μ sec. after the high-power transmitter in order to delay all low-power echoes by 20 km. Under normal, nonshower conditions the high-power system should detect $(2000)^{1/2}$ $(9.22/9.55)^{3/2} = 43$ times as many echoes as the low-power system (see next section for figure-of-merit formula).

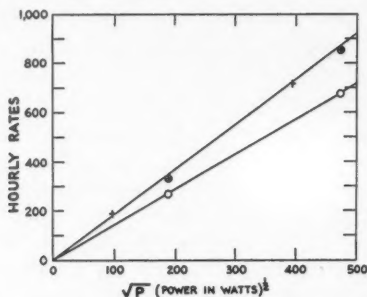


FIG. 3. Observed hourly rates of meteor echoes versus square root of transmitter power.
 Circled points—period 1900-0415 E.S.T., Oct. 3, 1950.
 Crosses—period 0000-0400 E.S.T., Nov. 29, 1950.
 Circled crosses—period 0000-0415 E.S.T., Oct. 3, 1950.

From 7.5 hr. of film records obtained between 1700 E.S.T. Jan. 8 and 0200 E.S.T. Jan. 9, 1951, the observed echo counts were as follows: high power 2287 echoes, low power 51 echoes; ratio 45 to 1. This is in good agreement with the predicted ratio for nonshower conditions.

On Jan. 3, 1951, the Quadrantid meteor shower was active. From 14 hr. of film records secured between 0900 E.S.T. and 2400 E.S.T. Jan. 3, 1951, the observed echo counts were as follows: high power 6973 echoes, low power 277 echoes; ratio 25.2 to 1. During the period 1200 E.S.T. to 1900 E.S.T., when the Quadrantid radiant was low in the sky, the high-low ratio averaged 19.6 to 1. As the radiant rose towards the "forbidden zone" (12) around the zenith fewer shower meteors were detected by either radar, and the high-low ratio increased. Comparing the shower and nonshower high-low ratios it follows that the low-power set detected more than twice as many echoes during the shower as would be predicted. The high-power hourly rates are themselves about twice the normal nonshower rates for the same hours in December and January, which would make the low-power hourly rates during the shower about four times the normal rates. The complete observations of the Quadrantid shower, including Doppler velocity measurements, will be presented elsewhere.

5. Variation of Rates with Wave Length

The variation of echo rates over a wide range of wave lengths is more difficult to evaluate than the variation with other parameters. It is not practical to attempt a rapid change of wave length of the radar system by a factor of two,

say, and if the observations are to be made simultaneously on two wave lengths, completely separate radar systems are required and the transmitter power, antenna coverage, transmission line losses, receiver band width and noise factor, and galactic noise should all be measured for each system. As it is usually difficult to make each of these parameters identical for the two systems (especially the galactic noise factor which varies widely with wave length and is not under the control of the experimenter) some method of putting the observations on a comparative basis must be used. If the same type of antenna is used on both wave lengths the evidence of the preceding sections suggests that the echo rates may be equated, as far as transmitter power and received signal power are concerned, by the use of a relative figure-of-merit, $F = (\epsilon_2 P_1 / \epsilon_1 P_2)^{1/2}$. Here ϵ_1 and ϵ_2 are the minimum detectable signal powers and are assumed to include factors such as receiver band width, receiver noise, and galactic noise. The minimum values of ϵ quoted in Table VI were determined

TABLE VI
COMPUTED FIGURE-OF-MERIT

Wave lengths (m.)	P (w.)	ϵ (watts)	F	$F \left(\frac{\lambda_1}{\lambda_2} \right)^{3/2}$	$F \left(\frac{\lambda_1}{\lambda_2} \right)^2$
9.22	2×10^6	2×10^{-14}	2.0	4.5	5.9
5.35	5×10^4	2×10^{-14}			
9.22	2×10^5	2×10^{-14}	8.2	48	87
2.83	3×10^4	2×10^{-13}			

under operational conditions by switching the receiver input alternatively between the antenna and a diode noise generator and equalizing the outputs on an A-scan oscilloscope. The measured values are about 6 db. or more above the minimum signal that could actually be read on the film record of the intensity-modulated display, owing to the integrating characteristics of the film. The measured values will be used here because they furnish reliable comparisons between receivers. This expression for F is consistent with both of the echoing formulae of Equations (1) and (2). The antenna gain factor has been left out of this discussion because it was expedient to use similar half-wave dipoles on all wave lengths, properly scaled in size and all oriented North-South. In this experiment it was necessary to measure the absolute power of each transmitter, hence the error in the relative power ratios was of the order of 20%.

The 9.22 m. radar system was first compared with a lower-powered equipment operating on 2.83 m. The short-wave system was later modified to operate on 5.35 m., primarily for some auroral research work. While the equipment was available advantage was taken of the opportunity to compare 9.22 m. with 5.35 m., using a half-wave dipole in each case. The minimum noise powers for

the 9.22 m. and 5.35 m. receivers were approximately equal, the decrease in galactic noise (14) at the shorter wave length being counterbalanced by the greater receiver band width. A new receiver was built for 5.35 m., but the 2.83 receiver was of early wartime construction and its performance inferior to similar receivers of modern design. The measurements are summarized in Table VI. In addition, the last two columns of Table VI show the figure-of-merit values multiplied respectively by the three-halves power and the square of the wave length ratios.

The observational results are presented in Table VII and Table VIII. Echoes were counted at all ranges up to the maximum of 260 km. available on the displays. Almost without exception every echo that occurred on the shorter wave lengths was detected on 9.22 m. The counts and the ratio are given for each hour to show the consistency of the data.

TABLE VII
OBSERVED ECHO RATIOS

Aug. 12, 1950 Hour E.S.T.	No. of echoes		Ratio
	9.22 m.	5.35 m.	
0000-0100	1,033	233	4.4
0100-0200	1,103	212	5.2
0200-0300	1,044	234	4.5
0300-0400	840	184	4.6
0400-0500	667	131	5.1
Mean	937	199	4.7

TABLE VIII
OBSERVED ECHO RATIOS

Apr. 3-4, 1950 Hour E.S.T.	No. of echoes		Ratio
	9.22 m.	2.83 m.	
2230-2330	514	5	103
2330-0030	565	7	81
0030-0130	707	7	101
0130-0230	691	7	99
0230-0330	654	4	164
Mean	626	6	104

From Table VII the mean ratio of echo counts, 4.7, agrees more closely with the predicted ratio of 4.5, based on Lovell's formula,* than with the value of 5.9 deduced from the assumption that the minimum echo strength varies as

* In a manuscript privately communicated Lovell proposed the use of a sensitivity factor, $S = G(P\lambda^2/\epsilon R^3)^{1/2}$, with $R = 100$ km. as a standard, for comparison of various equipments, but deleted this remark from the published paper (17). The quantity $F(\lambda_1/\lambda_2)^{3/2}$ used here expresses the same relation in relative form.

the fourth power of the wave length. Comparing the observed ratio in Table VIII with the theoretical ratios for this case the interpretation is indefinite as both theoretical predictions are below the observed value. A reason for the discrepancy will be suggested in the next section.

Referring to Table VII it should be remarked that the echo counts for the hours 0000-0300 E.S.T. include as many Perseid meteors as normal background meteors. By 0500 E.S.T. the Perseid radiant was so high in the sky that only the brighter Perseid meteors were being detected. The April period, Table VIII, was considered to be reasonably free of any pronounced meteor shower.

6. Confirmation of Lovell's Formula

We have examined statistically the possibility that the meteor trail may be a highly efficient filter-polarizer on the longer wave lengths and have rejected this hypothesis and its implied explanation of the unusually large numbers of meteor echoes reported on the longer wave lengths. Nor do we consider that multiple echoes from the brighter meteors can begin to account for the observed rates because the complex meteor echoes comprise only a small percentage of the total number of echoes observed on the longer wave lengths. Furthermore, such multiple echoes are usually clearly recognizable as being due to one meteor and are counted as one echo.

The observed variations of echo rates with transmitter power and antenna gain suggest a direct proportionality between echo rates and the minimum detectable value of α deduced from Lovell's formula, Equation (1b). These observations are equally in accord with Equation (2), based on a mechanism of total reflection from an ionized cloud with a finite scattering area. However the time factor suggests that Lovell's formula is to be preferred because most of the small echoes have been shown to occur simultaneously with the passage of the meteor across the line of sight (12). The diffraction echo from a line source is detectable without delay, whereas an appreciable time would elapse before the trail could expand to a volume having a cross-sectional area sufficient to yield an echo. Furthermore, the line densities of most of the detectable meteor trails are of the order 10^9 to 10^{10} electrons per cm. (assuming Lovell's formula) and it is likely that the volume density of the cloud would drop below the critical density before the minimum cross-sectional area was attained.

The observed variations of echo rates with wave length, though not as conclusive as the other observations, do indicate that the rates increase with wave length more rapidly than would be expected on the basis of Lovell's formula alone. The explanation may lie in the decreasing aspect sensitivity of the average meteor trail as the wave length is increased, an effect that was strikingly apparent in this work and which has been pointed out by Lovell. It has been shown (12) that the received power or intensity of the echo diffracted from the meteor trail as it is being produced increases proportionally to λ/s^2 , up to the final value ϵ given by Equation (1a), where s is the distance the meteor has still to travel along its path before reaching the t_0 point (the point

where the path is normal to the line of sight). Lovell (8) has put forward ample experimental verification that the maximum intensities of the echoes from a given meteor as observed on two wave lengths will vary as λ^3 . However, at a given distance prior to the t_0 point the ratio of the echo intensities will vary by an additional factor λ . In other words, the scattering or re-radiation pattern of the meteor trail is narrower on the shorter wave lengths, hence there is a smaller probability that a detectable echo will be received. On wave lengths in the range 20-100 m., say, the re-radiation patterns are much broader and the effective statistical variation of echo power with wave length would be as λ^3 multiplied by λ , that is λ^4 , hence the echo rates would vary as λ^2 .

7. Distribution of Nonshower Meteors

From a preliminary analysis of 1938 visual nonshower meteors observed between 2300 and 0100 E.S.T. on the clear moonless nights of the Ottawa meteor program, Millman has determined the hourly rates of meteors in various magnitude classes. In Fig. 4 each solid point represents the total number of meteors per hour, seen by a standard team of six observers, down to and including the magnitude given by the abscissa of the point, the data being available in half-magnitudes. The visual rates begin to fall off at the 2nd magnitude, for various physiological reasons. A solid curve is drawn through these observational points, terminating at the 5th magnitude which is the limit of the visual observers. The solid straight line, drawn through the linear part of the data, may be extrapolated to fainter magnitudes to represent the

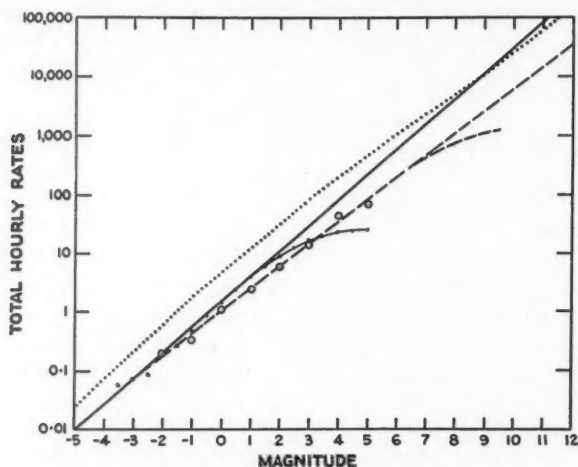


FIG. 4. Total hourly meteor rates versus limiting magnitude.
 Solid points—visual observations of nonshower meteors.
 Solid straight line—theoretical graph extrapolated from visual nonshower data.
 Circled points and dashed line—graph for 9.22 m. radar computed from visual-radar correlations.
 Dotted curve—proposed theoretical graph for shower meteors.

theoretical rates of available meteors. The slope of the solid line of Fig. 4 is 2.7. Watson (18) mentions the value 2.5 for the increase in numbers with each magnitude. If it is assumed that there is a constant ratio of numbers in any two consecutive magnitude classes then the ratio of the number of meteors of magnitude M_1 to the number of magnitude M_2 is the same as the ratio of the sum of all meteors from magnitude minus infinity down to M_1 to the sum of all meteors down to M_2 . The sums have been plotted in Fig. 4 for convenience, but a graph of numbers in each magnitude versus magnitude will have the same slope, though displaced downward by a factor 1.59 for the slope 2.7 and 1.67 for the slope 2.5.

The brightest meteors will be detected by the radar regardless of the path orientations. Because of the aspect effect the radar will miss an increasing proportion of meteors as the magnitudes become fainter. Assume that there are N meteors per hour of a given magnitude, theoretically available. Of this number, N , the visual observers record a known number n , and the radar detects an unknown number m . The sampling processes differ, of course, for the visual and radar meteors but for simplicity let us assume that each process is random. Suppose that p of the n visual meteors are definitely correlated with radar echoes, by methods outlined elsewhere (11). The proportion of radar meteors in the visual observations is then assumed to be the same as the proportion of all radar meteors to the total number available, i.e., $p/n = m/N$. For meteors brighter than 2nd magnitude, $n = N$. The ratios, p/n , can easily be determined up to the 5th magnitude, though the ratios will tend to be high because of the increasing probability of chance correlation as the fainter magnitude classes are considered. The circled points of Fig. 4 show the radar rates computed on this basis from 607 visual-radar correlations. The dashed straight line through the circled points of Fig. 4 indicates the tentative relationship between rates and magnitudes for the 9.22 m. radar system operating at a normal power of 200 kw.

Typical rates of 600–1000 radar meteors per hour at midnight on nonshower nights would correspond to a limiting magnitude of 7th to 8th magnitude for the radar system, read directly from the dashed line of Fig. 4. For several reasons this is to be regarded as an underestimate of the limiting sensitivity of the radar system. First, the computed points at 3rd to 5th magnitude on the radar graph will tend to be high. Second, if the slope of the theoretical visual line were changed from 2.7 to 2.5 the corresponding radar line would yield a limiting magnitude between the 8th and 9th. Third, if the radar rates were to fall off in a manner analogous to the visual, as suggested by the dashed curve of Fig. 4, the limiting magnitude would be even fainter. The reasons for such a decrease in the radar rates at the fainter magnitude would be quite different from the reasons for the visual rate decrease: one suggestion would be that radar meteors in the 9th or 10th magnitude ranges tend to occur at high elevations, on the upper fringe of the M-region (11), where less ionization is produced for a given mass and velocity of the particle. Fourth, an important consideration that has not yet been taken into account in the argument so far

is the fact that the visual data include some meteors beyond the range of the radar gear and many meteors near the limits of the radar range where the radar sensitivity is four or five magnitudes less than in the zenith. Magnitude corrections should be applied to the more distant visual meteors, as well, but these will be smaller than the radar corrections. Expressed in other words, the attenuation with range and zenith angle is less for the visual observers than for the radar. The average range of the radar echoes is about 130 km. where the sensitivity is about one magnitude less than in the zenith. Hence, the radar graph of Fig. 4 will apply to the average radar meteor, and the limit for overhead meteors will be one magnitude fainter.

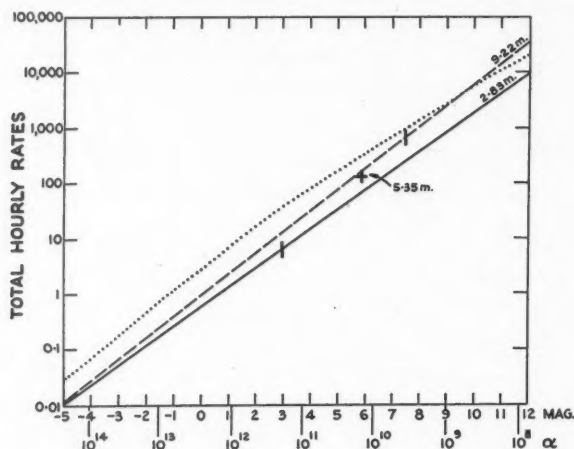


FIG. 5. Total hourly meteor rates versus limiting magnitude, and versus α , the number of electrons produced per cm. of meteor path.

Dashed line—graph for 9.22 m. radar (taken from Fig. 4).

Solid line—graph for 2.83 m. radar.

Dotted curve—proposed theoretical graph for shower meteors detected by the 9.22 m. radar.

The dashed line of Fig. 4 has been transferred to Fig. 5, and represents the uncorrected relation between normal, or nonshower radar echo rates and magnitudes for the 9.22 m. system. From Table VI the theoretical rate ratio for 9.22 versus 2.83 m. was computed to be about 50 to 1, whereas the observed ratio was about 100 to 1 (see Table VIII). As suggested above the discrepancy may be due in part to the effectively sharper radiation pattern of the meteor trail on the shorter wave length. This would imply that the short wave length radar system should lie on a graph of smaller slope than the long wave length system. From the 9.22 m. graph a rate of 600 per hour corresponds to a limiting sensitivity of 7.5 magnitude. If the power of the 9.22 m. radar had been lowered so that the rate was 12 per hour the limiting magnitude would be 3. However,

the observed rate using the 2.83 m. radar was 6 per hour, so that the graph for 2.83 m. should pass through this point on the ordinate at magnitude 3. We may assume that the short wave radar will see all of the extremely bright meteors, and draw the solid straight line, shown in Fig. 5, as indicative of the performance of a radar system on 2.83 m. The short, heavy vertical strokes on the graphs represent the minimum limiting magnitudes of the two equipments actually used. The 5.35 m. radar would be on a graph below but close to the 9.22 line. The limiting magnitude for the 5.35 m. system is estimated from Tables VI and VII to be about the 6th magnitude and is marked on Fig. 5 with a cross.

The predicted rate based on Lovell's published data of the parameters of his 4.2 m. radar system with a half-wave dipole antenna is in fair agreement with his observed nonshower rates, and suggest a limit of 5th magnitude for his equipment. Lovell (1) has stated that his gear has a limiting sensitivity of about 6th magnitude, presumably based on the visual scale used by Prentice and Porter. Clegg and Davidson (3) present data which suggest that the English visual observers record large numbers of 6th magnitude meteors, whereas our visual records contain none. If meteors described as 5th magnitude by the Canadian team were considered to be 6th magnitude meteors by the English observers then the magnitude scale of Figs. 4 and 5 should be shifted one magnitude to the left. Or, if a slope of 2.5 is used to represent the value generally quoted in the literature, instead of the tentative value 2.7 based on the observations made at Ottawa, a similar displacement of the limiting sensitivities towards the fainter magnitudes will be obtained, although the displacement will be greater for the fainter magnitudes. To sum up, it appears that the limiting magnitude of the 9.22 m. radar, operating at 200 kw. power into a half-wave dipole, is certainly fainter than the 8th magnitude and is probably of the order of 9th to 10th magnitude.

The limiting magnitude of the continuous-wave Doppler system employed at Ottawa for the measurement of meteor velocities (12) may be deduced on a relative basis. The Doppler wave length is 9.99 m., the transmitter power 2 kw. and the receiver band width 2 kc. per sec. The band width of the radar receiver is 200 kc. per sec. Theoretically, the minimum noise powers of the two receivers should vary as the band width ratio, or $\epsilon_{\text{radar}}/\epsilon_{\text{Doppler}} = 100$. This ratio exactly counterbalances the transmitter power ratio, hence the relative figure-of-merit of Doppler versus radar should depend only on the wave length ratio $(9.99/9.22)^{3/2} = 1.13$. The Doppler should thus be slightly more sensitive than the radar, but the actual Doppler echo counts are found to vary from 1/10 to 1/3 of the radar echo counts. The discrepancy between theory and practice is explained in part by an additional source of noise being present in the Doppler system, owing to scattered transmitter radiation of random phase reaching the receiver by reflections from ground objects in the vicinity. Tree branches waving in the wind can increase the Doppler receiver noise figure significantly. This effect does not increase the radar noise figure,

except for the first few kilometers of the range sweep where no meteor echoes occur. Using the experimental data rather than the theoretical values, it follows that the Doppler sensitivity is less than that of the radar by one or two magnitudes, depending on local conditions.

8. Distribution of Shower Meteors

Evidence was put forward in a previous paper (11) which suggested that the numbers/magnitude relation for a strong visual shower does not follow the same law as for the normal background meteors. The maximum ratio of shower echoes to background echoes during the Geminid shower of 1947 was about 1 to 1 whereas the visual ratio was between 2 to 1 and 3 to 1. It was also pointed out that the percentage of radar echoes of duration 1 sec. and longer during the Geminid shower was twice as great as the percentage during typical nonshower periods. The conclusion was that there is a smaller percentage of faint meteors in the shower stream than in the normal background, the implication being that some agency, such as the Poynting-Robertson effect, has weeded out the smaller shower meteors over the course of centuries. The evidence of the high-low power experiments with the 100-watt and 200 kw. radar system showed that the echo rates obtained during the Quadrantid shower with the low-power transmitter were twice the rates that would be expected if the distribution of the shower meteors were to follow the normal distribution law. From the predicted nonshower ratio, 43 to 1, or the observed ratio, 45 to 1, the limiting magnitude of the 100 w. set is estimated from Fig. 5 to be about 3rd or 4th magnitude, which is well within the visual range.

Let us return to Fig. 4 and draw the dotted curve as a tentative representation of the theoretical numbers/magnitude distribution of shower meteors in a well established shower such as the Perseids or Geminids. In the range -3 to $+2$ magnitude, the indicated shower rates are three times the normal rates, but at fainter magnitudes the shower curve is arbitrarily drawn to approach the normal line and to cross it at about the 9th magnitude. Over the visual range the ratio of shower to nonshower meteors would be of the order of 3 to 1, though there might be a slight decrease noticed at 5th or 6th magnitude. However, at 9th magnitude the ratio would be about 1 to 1. A similar shower curve for radar meteors is drawn as the dotted line in Fig. 5, again crossing over at about the 9th magnitude. The radar rates of the 200 kw. 9.22 m. equipment would then have doubled during a strong shower while the visual rates would have quadrupled, in qualitative agreement with observation. Radar equipment having a lower sensitivity, whether by virtue of lower power or shorter wave length, would record an increase in echo rates during showers more in proportion to the increase in visual meteor rates. Lovell has stated that his equipment on 4.2 m. does indeed yield echo rates in proportion to the visual rates.

Olivier (15) has concluded from his studies of telescopic meteors that the meteor rates begin to depart from the expected law of increase of number at

the 7th or 8th magnitude, and to fall off rapidly after the 9th or 10th magnitude. He does not state explicitly what fraction of his observations might be shower meteors: if the proportion were large his findings would be in qualitative agreement with our own. However, if most of his observations are nonshower meteors it would imply that our theoretical numbers/magnitude relation for normal nonshower meteors (solid line, Fig. 4), and also our normal radar graph (dashed line, Fig. 4 or Fig. 5), should commence to curve downwards at the 7th or 8th magnitude. If this were so the limiting sensitivity of the high-power radar would be even fainter than estimated above, possibly of the order of 12th magnitude, but the high-low power observations do not lend support to the suggestion.

It is of interest to speculate upon the performance of a 10 kw. radar set operating at 100 m., say. Despite the lower power the echo rates should be increased by a factor of 8 over the 9.22 m. gear. The 100 m. set would be operating on a line above that for the 9.22 m. set and below the theoretical graph of available meteors. Roughly speaking, the sensitivity of the low-power 100 m. set would be more than two magnitudes fainter than the high-power 9.22 m. set, and the over-all meteor rates seen during well known showers by the long wave set would show little or no increase over normal rates, although the percentage of long-enduring echoes would increase.

On the other hand, if the earth were to pass through clouds of meteoric dust one might expect very great numbers of extremely faint meteors to occur, which would escape detection by high-power radars on 10 m. or shorter wave lengths but would be seen by long wave equipments. The echo rates would be so high that the resolution of the conventional ionospheric display would be inadequate to show individual echoes, and a continuous band echo should be obtained. This is purely speculative but it may offer an explanation for some sporadic-E type echoes detected by ionosphere equipments of moderate power working in the range 50-150 m. If any sporadic-E type echoes were due to these dust clouds the echoes should appear simultaneously over a hemisphere of the earth. Pineo (16) has indicated that sporadic-E echoes in general cannot be correlated with the well established visual showers. Our own observational data confirm his conclusion, which is compatible with the argument advanced above that the visual showers should have little effect on ionosphere equipment. The suggestion that showers of faint meteors can occur which would affect ionospheric records may be worth some investigation. These would not necessarily be meteor showers having predictable orbits in the usual sense. They would more likely be due to chance encounter with relatively dense clouds of dust, possibly interstellar material.

Using Equation 1(b) it is found that the minimum line density of electrons that can be detected by the 9.22 m. radar is of the order of 10^9 to 5×10^9 electrons per cm. We may associate the value $\alpha = 10^9$ with the 9th magnitude, estimated above as the limit for the 9.22 m. set. It was shown that meteor rates varied as the square root of the power, or directly with α , hence the

complete correlation between α and visual magnitude may easily be established. This relation is shown as the second abscissa scale of Fig. 5. An independent line of reasoning led Herlofson (5) to estimate that $\alpha = 10^{12}$ would correspond to a 1st magnitude meteor and $\alpha = 10^{10}$ to a 6th magnitude meteor, which is in agreement with the scale of Fig. 5.

Summing up, it has been demonstrated that Lovell's formula is in accord with the experimental evidence over a wave length range of 2.8 to 10 m., when consideration is given to the effective broadening of the radiation pattern of the meteor trail as the wave length is increased. No difficulty is encountered in explaining the higher meteor rates observed on the longer wave lengths, and there appears to be no need to postulate additional echoing mechanisms. In particular, the filter-polarizer mechanism based on Herlofson's plasma theory has received no support. The increase of visual rates over radar rates during a strong visual shower, compared to the rates observed during normal or nonshower periods, can be explained by assuming that there are fewer small meteors in the shower stream. The argument has tentatively been extended to wave lengths of 50 to 100 m. to suggest that echoes might be obtained from clouds of meteoric dust which could be detected only by long wave radars, with the appearance of a sporadic-E type echo.

Acknowledgments

At one time or another every member of the Radiophysics Section of the Division of Radio and Electrical Engineering has been associated with these experiments. Particular acknowledgment should be made to Mr. E. L. R. Webb and Mr. B. E. Bourne who designed most of the equipment, and to Mr. S. M. Panagapko who reduced the radar film records. Dr. Peter M. Millman and his staff of the Dominion Observatory supplied the visual meteor data. In the preparation of this paper the advice and assistance of Dr. Barbara M. McKinley and Dr. Millman have been keenly appreciated.

References

1. ALMOND, MARY, DAVIES, J. G., and LOVELL, A. C. B. *Observatory*, 70: 112. 1950.
2. BLACKETT, P. M. S. and LOVELL, A. C. B. *Proc. Roy. Soc. (London) A*, 177: 183. 1941.
3. CLEGG, J. A. and DAVIDSON, I. A. *Phil. Mag. Ser. 7*, 41: 77. 1950.
4. HERLOFSON, N. *Observatory*, 68: 226. 1948.
5. HERLOFSON, N. *Reports on Progress in Physics*, 11: 444. 1948.
6. HEY, J. S. and STEWART, G. S. *Proc. Phys. Soc. (London)* 59: 858. 1947.
7. LOVELL, A. C. B., BANWELL, C. J., and CLEGG, J. A. *Monthly Notices Roy. Astron. Soc.* 107: 164. 1947.
8. LOVELL, A. C. B. and CLEGG, J. A. *Proc. Phys. Soc. (London)* 60: 491. 1948.
9. LOVELL, A. C. B. *Reports on Progress in Physics*, 1: 415. 1948.
10. LOVELL, A. C. B. *Science Progress*, 38: 22. 1950.
11. MCKINLEY, D. W. R. and MILLMAN, PETER M. *Proc. I.R.E.* 37: 364. 1949.
12. MCKINLEY, D. W. R. *Astrophys. J.* 113: 225. 1951.
13. MILLMAN, PETER M. and MCKINLEY, D. W. R. *Sky and Telescope*, 8: 114. 1949.
14. NORTON, K. A. and OMBERG, A. C. *Proc. I.R.E.* 35: 4. 1947.
15. OLIVIER, CHARLES P. *Proc. Am. Phil. Soc.* 94: 327. 1950.
16. PINEO, V. C. *Science*, 112: 50. 1950.
17. PRENTICE, J. P. M., LOVELL, A. C. B., and BANWELL, C. J. *Monthly Notices Roy. Astron. Soc.* 107: 155. 1947.
18. WATSON, FLETCHER G. *Between the planets*. 1st Edition. The Blakiston Co., Philadelphia. 1941. p. 114.

A UNIFIED PICTURE OF DIFFUSION¹

By J. D. BABBITT

Abstract

It is shown that if the equation

$$\frac{\partial p}{\partial x} - A_x = 0,$$

where p is a pressure function and A_x a resistive or frictional force, is taken as the fundamental dynamical equation, it is possible to explain diffusion in solutions, in solid solutions, and in an adsorbing solid as well as the interdiffusion of two gases. In diffusion in both liquid and solid solutions the pressure function is the osmotic pressure; in diffusion of adsorbed gases the pressure function is the spreading pressure of the adsorbed layer. Since these pressure functions are thermodynamic properties of the system the pressure gradient is independent of the physical picture assumed for the system. To evaluate A_x , however, it is necessary to assume a specific physical mechanism for the movement of the molecules.

Recently (1) we have shown that if, in place of Fick's law, giving the flux of diffusion as proportional to the concentration gradient, it is assumed that a pressure gradient is the operative force for diffusion, then it is possible to obtain a reasonable picture of the diffusion of gases through solids. At that time we were mainly interested in finding some explanation of the flow of adsorbed gases through adsorbent solids. Since that paper was written, we have become aware, through a recent publication (4), of the work that has been done on the diffusion of metals in metals and of the theories that have been put forward to explain such diffusion. After consideration of this work, it became apparent that an extension of the ideas presented earlier would lead to the same equations as those adopted to explain the diffusion of metals and would, in consequence, present a single picture covering all aspects of diffusion. We shall, then, in this paper, apply the fundamental equation developed earlier to diffusion in liquid and solid solutions and show that this equation leads to the same expression as the current theories of such diffusion.

Fundamental Equation and Application to Gases

In the earlier work it was shown that to derive the differential equation for diffusion it was necessary to combine

- (i) a dynamical equation defining the motion,
- (ii) an equation of continuity expressing the law of conservation of mass,
- and (iii) an equation of state.

In the past Fick's law

$$P = -D_c \frac{\partial c}{\partial x} \quad (1)$$

¹ Manuscript received September 18, 1950.

Contribution from the Division of Physics, National Research Laboratories, Ottawa, Canada. Issued as N.R.C. No. 2509.

has been implicitly used as the fundamental dynamical equation. Here \mathbf{P} is the rate of permeation or the flux diffusing, D_c is the diffusion coefficient and c is the concentration. From hydrodynamical considerations the fundamental equation should have the form

$$\frac{\partial p}{\partial x} - A_x = 0, \quad (2)$$

where p is a potential function having the properties of a pressure and A_x is the resistive or frictional force that opposes the motion. This equation has general validity for diffusion when inertial effects and external forces can be neglected whereas Fick's law is a secondary law derived from it under the special conditions arising when c is directly proportional to p . To apply Equation (2) to the various types of diffusion it is necessary only to choose correctly the pressure function that is operative and to express the resistive forces in terms of the physical properties of the system.

It was shown long ago by Maxwell (5) that the application of Equation (2) to the interdiffusion of two gases presents no difficulty and, in fact, leads directly to Fick's law when the gases obey the gas law. Maxwell's fundamental equations for the interdiffusion of two gases are

$$\frac{\partial p_1}{\partial x} + C\rho_1\rho_2(u_1 - u_2) = 0 \quad \text{and} \quad \frac{\partial p_2}{\partial x} + C\rho_1\rho_2(u_2 - u_1) = 0 \quad (3)$$

where p_1, p_2 , ρ_1, ρ_2 , and u_1, u_2 are the partial pressures, the densities, and the velocities of the two gases A and B respectively and C is the coefficient of resistance. C represents the force that must be applied to unit mass of the gas to maintain unit velocity when the density of the second gas is unity.

It is easy to prove that

$$u_1 - u_2 = \frac{u_1 P}{p_2} = -\frac{u_2 P}{p_1}, \quad (4)$$

where $P = p_1 + p_2$ = total pressure. Thus we obtain

$$\rho_1 u_1 = -\frac{p_2}{C\rho_2 P} \frac{\partial p_1}{\partial x} \quad \text{and} \quad \rho_2 u_2 = -\frac{p_1}{C\rho_1 P} \frac{\partial p_2}{\partial x}. \quad (5)$$

If we assume that the gases obey the gas law so that

$$p_1 = \frac{RT}{M_1} \rho_1 \quad \text{and} \quad p_2 = \frac{RT}{M_2} \rho_2,$$

$$\text{then} \quad \rho_1 u_1 = -\frac{(RT)^2}{CPM_1 M_2} \frac{\partial \rho_1}{\partial x} \quad \text{and} \quad \rho_2 u_2 = -\frac{(RT)^2}{CPM_1 M_2} \frac{\partial \rho_2}{\partial x}. \quad (6)$$

If we place $\frac{(RT)^2}{CPM_1 M_2} = D$, Equations (6) are equivalent to Fick's law. Thus,

if the gases behave as perfect gases, Fick's law is a valid expression for the diffusion but when the laws of perfect gases are not applicable it is necessary to go to Equations (5) for the fundamental equation.

Diffusion in Solutions

We may apply the same concepts that were used above for the interdiffusion of two gases to the diffusion of a solute, B , into a solvent, A . As before we start with the fundamental equation

$$\frac{\partial p}{\partial x} - A_x = 0$$

In the theory of solutions it has long been known that the osmotic pressure is a measure of the tendency of a solute to dilute itself and as early as 1888 Nernst (6) assumed that diffusion in a solution was proportional to the gradient of the osmotic pressure. A similar approach has been used by Sackur (8), and Bryan and Porter in their article on diffusion in the 14th edition of the *Encyclopaedia Britannica* also suggest that in solutions the osmotic pressure is the operative force for diffusion. It is, therefore, not only a reasonable assumption but one that has considerable support to assume that the osmotic pressure is the pressure function that should be used for solutions in Equation (2). Before replacing the partial pressures of gaseous diffusion with the osmotic pressures of solutions it is necessary, however, to consider carefully the way in which the term osmotic pressure is used.

The osmotic pressure has been defined by Guggenheim (2) as follows. If a solvent and solution are separated by a membrane that allows the solvent to pass without hindrance, but prevents the passage of the dissolved substance, the solvent will tend to flow through the membrane into the solution. The hydrostatic pressure that must be applied to the solution to stop the flow of solvent through the membrane is the osmotic pressure of the solvent. It is the difference of the pressure of the solvent in the pure phase and in the solution. It is obvious that the pressure tending to make the solvent molecules diffuse is equal to the osmotic pressure. It is also apparent that, if we could find a semi-permeable membrane that would allow the passage of the solute molecules but prevent the passage of the solvent molecules, exactly the same considerations would hold and the osmotic pressure of the solute atoms would be a measure of the tendency of the solute molecules to diffuse. In this case the pressure responsible for the diffusion would be equal to the pressure of solute molecules in solution since their pressure in the pure solvent would be nil.

In this picture, the osmotic pressure is defined as the hydrostatic pressure that must be applied to stop diffusion; it follows, therefore, that the pressure causing the diffusion would be equal but opposite in sign to the osmotic pressure. It is interesting to note that, since the concept of a semipermeable membrane could be applied equally well to a mixture of gases, the partial pressures would play there the role assigned in solutions to the osmotic pressures.

To determine A_x we proceed as for gases and assume that if $\rho_1 \rho_2$ are the mass of solute and solvent respectively per cc., u_1 and u_2 their velocities and C the

coefficient of resistance, then the portion of the solvent contained in a small element of volume v will experience from the solute a resistance $C\rho_1\rho_2v(u_1 - u_2)$ and the solute contained in the same volume element will experience from the solvent an equal and opposite resistance $C\rho_1\rho_2v(u_2 - u_1)$. With these considerations we write for the equations of flow

$$\frac{\partial \pi_1}{\partial x} + C\rho_1\rho_2(u_1 - u_2) = 0 \quad \text{and} \quad \frac{\partial \pi_2}{\partial x} + C\rho_1\rho_2(u_2 - u_1) = 0 \quad (7)$$

where π is the osmotic pressure defined above. The positive sign is used since the resistive force is in the same direction as the increase of pressure.

In a solution the hydrostatic pressure will remain constant so that the net transfer of mass across a plane must be zero.

$$\text{Therefore} \quad \rho_1 u_1 + \rho_2 u_2 = 0. \quad (8)$$

$$\text{Hence} \quad \frac{u_1}{\rho_2} = -\frac{u_2}{\rho_1} = \frac{u_1 - u_2}{\rho_1 + \rho_2} = \frac{u_1 - u_2}{d}, \quad (9)$$

where d is the total mass per cc., that is, the density of the solution.

Substituting in (7), we get

$$\frac{\partial \pi_1}{\partial x} + Cd\rho_1 u_1 = 0 \quad \text{and} \quad \frac{\partial \pi_2}{\partial x} + Cd\rho_2 u_2 = 0. \quad (10)$$

Transposing we obtain our equation of diffusion giving the mass flux diffusing across a plane as

$$\rho_1 u_1 = -\frac{1}{Cd} \frac{\partial \pi_1}{\partial x} \quad \text{and} \quad \rho_2 u_2 = -\frac{1}{Cd} \frac{\partial \pi_2}{\partial x}. \quad (11)$$

Now if van't Hoff's law, $\pi = RTc$, where c is concentration in gm-mol. per cc., holds then we can write

$$\frac{\partial \pi}{\partial x} = \frac{RT}{M} \frac{\partial \rho}{\partial x} \quad (12)$$

and our equations become

$$\rho_1 u_1 = -\frac{RT}{CdM_1} \frac{\partial \rho_1}{\partial x} \quad \text{and} \quad \rho_2 u_2 = -\frac{RT}{CdM_2} \frac{\partial \rho_2}{\partial x}. \quad (13)$$

Thus we see that, for solutions that obey van't Hoff's law, Fick's law is valid. It is to be noted, however, that Equations (11) are generally applicable, whereas Equations (13) are restricted to ideal solutions. To find a general relation between diffusion and concentration gradient it is necessary to introduce the relation between π and the concentration for nonideal solutions and that we shall now proceed to do.

Equation of Diffusion in Terms of Concentration Gradient

According to Guggenheim (2) the osmotic pressure of the solvent in a non-ideal solution is given by

$$\pi_1 = \frac{RT}{[V_1]} \log \frac{1}{N_1 f_1} \quad (14)$$

where $[V_1]$ is the partial molar volume of the solvent, N_1 is the partial molar fraction, and f_1 is the activity coefficient. It must be remembered that the activating pressure for diffusion is opposite and equal to this. Therefore we can write

$$\pi_1 = \frac{RT}{[V_1]} \log N_1 f_1 \quad \text{and} \quad \pi_2 = \frac{RT}{[V_2]} \log N_2 f_2. \quad (15)$$

Here $[V_2]$ is the partial molar volume of solute.

The chemical potentials of the two substances are given by

$$\mu_1 = \mu_1^\circ + RT \log N_1 f_1 \quad \text{and} \quad \mu_2 = \mu_2^\circ + RT \log N_2 f_2 \quad (16)$$

where μ_1° and μ_2° are constants depending only on temperature and pressure.

From (15) and (16) we have

$$\frac{\partial \pi_1}{\partial x} = \frac{1}{[V_1]} \frac{\partial \mu_1}{\partial x} \quad \text{and} \quad \frac{\partial \pi_2}{\partial x} = \frac{1}{[V_2]} \frac{\partial \mu_2}{\partial x}. \quad (17)$$

Substituting in Equation (11) we get

$$\rho_1 u_1 = - \frac{1}{Cd[V_1]} \frac{\partial \mu_1}{\partial x} \quad \text{and} \quad \rho_2 u_2 = - \frac{1}{Cd[V_2]} \frac{\partial \mu_2}{\partial x} \quad (18)$$

Putting $\frac{1}{Cd[V_1]} \equiv \Omega_1$ and $\frac{1}{Cd[V_2]} \equiv \Omega_2$ we may write these equations as

$$\rho_1 u_1 = - \Omega_1 \frac{\partial \mu_1}{\partial x} \quad \text{and} \quad \rho_2 u_2 = - \Omega_2 \frac{\partial \mu_2}{\partial x} \quad (19)$$

We see, therefore, that the mass flux of diffusion is proportional to the gradient of the chemical potential.

Several workers have suggested that diffusion should be taken as proportional to the gradient of chemical potential rather than to the concentration gradient (*cf.* Onsager and Fuoss (7), Jost (3), LeClaire (4)). Although this move seems to have been made intuitively by these workers, the above analysis shows that it is a reasonable deduction from first principles. Proceeding, we have from (16)

$$\frac{\partial \mu_1}{\partial x} = RT \frac{\partial}{\partial x} \log N_1 f_1 = RT \frac{\partial}{\partial x} \log \frac{n_1 f_1}{N} \quad (20)$$

where n_1 = number of solvent molecules and N total number of molecules per cc. Hence

$$\frac{\partial \mu_1}{\partial x} = RT \frac{\partial}{\partial x} [\log n_1 + \log f_1 - \log N].$$

We may consider N as constant so that

$$\frac{\partial \mu_1}{\partial x} = RT \left[\frac{1}{n_1} + \frac{\partial \log f_1}{\partial n_1} \right] \frac{\partial n_1}{\partial x}, \quad (21)$$

$$\begin{aligned} &= \frac{RT}{n_1} \left[1 + \frac{\partial \log f_1}{\partial \log n_1} \right] \frac{\partial n_1}{\partial x}, \\ &= \frac{RT}{\rho_1} \left[1 + \frac{\partial \log f_1}{\partial \log n_1} \right] \frac{\partial \rho_1}{\partial x}. \end{aligned} \quad (22)$$

Similarly

$$\frac{\partial \mu_2}{\partial x} = \frac{RT}{\rho_2} \left[1 + \frac{\partial \log f_2}{\partial \log n_2} \right] \frac{\partial \rho_2}{\partial x}.$$

Substituting in (18) we get

$$\begin{aligned} \rho_1 u_1 &= - \frac{RT}{Cd\rho_1[V_1]} \left[1 + \frac{\partial \log f_1}{\partial \log n_1} \right] \frac{\partial \rho_1}{\partial x} \quad \text{and} \\ \rho_2 u_2 &= - \frac{RT}{Cd\rho_2[V_2]} \left[1 + \frac{\partial \log f_2}{\partial \log n_2} \right] \frac{\partial \rho_2}{\partial x}, \end{aligned} \quad (23)$$

$$\begin{aligned} \rho_1 u_1 &= - \frac{\Omega_1 RT}{\rho_1} \left[1 + \frac{\partial \log f_1}{\partial \log n_1} \right] \frac{\partial \rho_1}{\partial x} \quad \text{and} \\ \rho_2 u_2 &= - \frac{\Omega_2 RT}{\rho_2} \left[1 + \frac{\partial \log f_2}{\partial \log n_2} \right] \frac{\partial \rho_2}{\partial x}. \end{aligned} \quad (24)$$

These equations are equivalent to those given by Onsager and Fuoss (7).

Introduction of the Mobilities

Since in the past it has been customary to use the concept of mobility in treating diffusion in solutions, it is interesting to introduce the mobility into our equations. We define the mobility G as the velocity of flow under unit force. Since our fundamental assumption is that the gradient of osmotic pressure is the force on the particles, the mobility is obtained by putting $\frac{\partial \pi}{\partial x} = 1$ in Equations (11).

$$\text{Thus} \quad G_1 = - \frac{1}{Cd\rho_1} \quad \text{and} \quad G_2 = - \frac{1}{Cd\rho_2}. \quad (25)$$

Substituting in (23) we obtain our equations of diffusion in terms of the mobilities

$$\rho_1 u_1 = \frac{G_1 RT}{[V_1]} \left[1 + \frac{\partial \log f_1}{\partial \log n_1} \right] \frac{\partial \rho_1}{\partial x} \quad \text{and} \quad \rho_2 u_2 = \frac{G_2 RT}{[V_2]} \left[1 + \frac{\partial \log f_2}{\partial \log n_2} \right] \frac{\partial \rho_2}{\partial x}. \quad (26)$$

In obtaining these formulas we have defined the mobilities according to the picture of diffusion that has been developed here. This, of course, is not the procedure that has been used hitherto and, for comparison, it is interesting to investigate the form which the equations would take if we had assumed, as has

been done in the past, that the driving force for the diffusion was the gradient of chemical potential rather than the osmotic pressure gradient.

Putting $\frac{\partial \mu}{\partial x} = 1$ in (18) we obtain for G

$$G_1 = -\frac{1}{Cd\rho_1[V_1]} = -\frac{\Omega_1}{\rho_1} \quad \text{and} \quad G_2 = -\frac{1}{Cd\rho_2[V_2]} = -\frac{\Omega_2}{\rho_2}. \quad (27)$$

This value for G is similar to that used by Onsager and Fuoss (7).

Proceeding on the assumption that G is defined by (27) we should get

$$\rho_1 u_1 = G_1 RT \left[1 + \frac{\partial \log f_1}{\partial \log n_1} \right] \frac{\partial \rho_1}{\partial x} \quad \text{and} \quad \rho_2 u_2 = G_2 RT \left[1 + \frac{\partial \log f_2}{\partial \log n_2} \right] \frac{\partial \rho_2}{\partial x}, \quad (28)$$

which is the equation given by Onsager and Fuoss (7) in terms of the mobility.

Equations (26) and (28) differ in the factor $\frac{1}{[V]}$. To complete the picture, it is necessary to see where this factor arises. Clearly it is the result of the two ways in which G has been defined—in one picture the force has been taken as equal to $\frac{\partial \pi}{\partial x}$ and in the other as equal to $\frac{\partial \mu}{\partial x}$. It is axiomatic that a spatial pressure gradient represents a force tending towards expansion but the meaning of a spatial gradient of chemical potential is not immediately evident. The osmotic pressure is a measure of the force tending to make one constituent of a solution change its volume whereas the chemical potential is a measure of the force tending towards a change of concentration. A change in volume will result in a change of concentration and Guggenheim (2) has pointed out that n bears to μ the same relation that v bears to p . Therefore it is reasonable to assume that $\frac{\partial \mu}{\partial n}$ is equivalent to $\frac{\partial p}{\partial v}$. Since, except for a constant factor, $\frac{\partial p}{\partial v}$ is equal to $\frac{\partial p}{\partial x}$ in one dimensional flow, and since $\frac{\partial p}{\partial x}$ is quite obviously the force behind any diffusion, it follows that $\frac{\partial \mu}{\partial n}$ rather than $\frac{\partial \mu}{\partial x}$ is the true representation in terms of the chemical potential of the force causing diffusion. Thus the two equivalent equations are

$$\rho u = D \frac{\partial p}{\partial v} \quad \text{and} \quad \rho u = D \frac{\partial \mu}{\partial n}.$$

That these are equivalent is immediately evident from the fundamental thermodynamic relation, $\frac{\partial \mu}{\partial p} \cdot \frac{\partial v}{\partial n} = 1$.

The origin of the factor $\frac{1}{[V]}$ is now obvious. We have

$$\begin{aligned}
 \rho u &= D \frac{\partial \mu}{\partial n}, \\
 &= D \frac{\partial v}{\partial n} \cdot \frac{\partial \mu}{\partial v}, \\
 &= D [V] \frac{\partial \mu}{\partial v},
 \end{aligned}$$

since $\frac{\partial v}{\partial n}$ is by definition the partial molar volume. Since again in one dimensional flow the change of independent variable from v to x involves only a constant factor our result is obtained.

Diffusion of a Solute into a Solvent Assuming that the Solvent is Stationary

In deriving these equations for diffusion in solutions we have assumed that both solvent and solute diffuse or, what is equivalent, that there is no distinction between solvent and solute. In dilute solutions a negligible error is introduced by assuming that the solvent is present in a much greater quantity than the solute and that it may be considered as stationary so that the diffusion is confined to the solute particles. Since the resulting equations are simpler than those produced above and have considerable application it is worthwhile to derive them.

As before we assume that $\frac{\partial \pi}{\partial x}$ is the driving force for the diffusion but π is now the osmotic pressure of the solute. The resistive force A_x may be written $C\rho u$ where C now includes a term for the density of the solvent; ρ is the concentration of solute per cc. and u is the velocity of solute with reference to the solvent. Thus

$$\frac{\partial \pi}{\partial x} + C\rho u = 0 \quad (29)$$

$$\text{and} \quad \rho u = -\frac{1}{C} \frac{\partial \pi}{\partial x}. \quad (30)$$

When van't Hoff's law holds $\frac{\partial \pi}{\partial x} = \frac{RT}{M} \frac{\partial \rho}{\partial x}$ as before and the equation becomes

$$\rho u = -\frac{RT}{CM} \frac{\partial \rho}{\partial x}. \quad (31)$$

In terms of the chemical potentials the same considerations as above hold so that

$$\rho u = -\frac{1}{C[V]} \frac{\partial \mu}{\partial x} \quad (32)$$

and the equation in terms of the concentration gradient becomes for nonideal solutions

$$\rho u = -\frac{RT}{C\rho[V]}\left[1 + \frac{\partial \log f}{\partial \log n}\right]\frac{\partial \rho}{\partial x} \quad (33)$$

The mobility G is now given by

$$G = -\frac{1}{C\rho}$$

so that G is the reciprocal of the resistance per cc.

The diffusion equation for nonideal solutions in terms of the mobility is

$$\rho u = \frac{GRT}{[V]}\left[1 + \frac{\partial \log f}{\partial \log n}\right]\frac{\partial \rho}{\partial x} \quad (35)$$

Diffusion of Metals in Metals

It is obvious that if we have a solid solution of one metal in another, we can apply, with very few changes, all the considerations given above for liquid solutions. This is, in effect, what has been done in the recent treatment of diffusion in metals which has been so ably reviewed by Le Claire (4). It is true that in solid solutions the concept of osmotic pressure has not been developed but there can be no doubt that if the molecules of the solute metal are free to change their positions in the solvent metal they will exert a pressure similar to osmotic pressure and this pressure will be a direct measure of the tendency for diffusion. Such a pressure can be evaluated thermodynamically. In metals, however, the chemical potential is a quantity that is more readily available, and, in consequence, it is probably preferable to express the diffusion as a function of the gradient of chemical potential. Since the gradients of chemical potential and osmotic pressure differ only in the factor $\frac{1}{[V]}$ it is immaterial which is used. The concept of osmotic pressure is useful, however, in relating the diffusion coefficient to the general hydrodynamic equation of flow where the resistive forces are balanced by a pressure gradient.

Once a fundamental dynamical equation for the diffusion in metals is obtained, the point of great interest is to relate the coefficient of resistance C or, what is equivalent, the mobility G , to the physical properties of the metal and specifically to the different models of atomic migration. In this paper, however, we are not concerned with that problem and shall proceed no further with it.

Diffusion of Gases through Adsorbing Solids

In an earlier paper (1) we have shown that Equation (2) may be applied to the diffusion of adsorbed gases through solids if the spreading pressure of the adsorbed layer is taken as the operative force for the diffusion. This idea follows logically from the unified picture presented here since the spreading pressure of adsorption is thermodynamically equivalent to the osmotic pres-

sure of solutions. Since the pressure may be obtained from the thermodynamic relation

$$p = - \left(\frac{\partial F}{\partial V} \right)_{T,n}, \quad (36)$$

where F is the free energy, it matters not, in determining the pressure gradient, whether the relation between the gas and the solid be assumed to be that of an adsorption or of a solution. To realize this is to eliminate immediately much of the confusion that has existed in the past in all theories of diffusion. In evaluating A_x , however, it is necessary to assume some physical picture for the diffusion since only through such a picture is it possible to relate the resistive force to the physical properties of the system.

References

1. BABBITT, J. D. *Can. J. Research, A*, 28: 449. 1950.
2. GUGGENHEIM, E. A. *Modern thermodynamics*. Methuen & Co. Ltd., London. 1933.
3. JOST, W. *Diffusion und chemische Reaktion in festen Stoffen*. Theodor Steinkopff. Dresden and Leipzig. 1937.
4. LECLAIRE, A. D. *Progress in metal physics*. Vol. 1. Chapter VII. *Edited by Bruce Chalmers*. Butterworth's Scientific Publications. 1949.
5. MAXWELL, C. *The scientific papers of Clerk Maxwell*. Vol. 2. Cambridge University Press. 1890. p. 623.
6. NERNST, W. *Z. physik. Chem.* 2: 613. 1888.
7. ONSAGER, L. and FUOSS, R. M. *J. Phys. Chem.* 36: 2689. 1932.
8. SACKUR, O. *Thermochemistry and thermodynamics*. *Translated by G. E. Gibson*. MacMillan & Co. Toronto. 1917. p. 276.

ON THE DIFFUSION OF ADSORBED GASES THROUGH SOLIDS¹

By J. D. BABBITT

Abstract

The equations for the diffusion of an adsorbed gas through a solid are developed on the assumption that the resistive force is directly proportional to the number of molecules adsorbed, that is, that the resistive force per molecule is constant. The equations are compared with experiment and also with the equations developed in an earlier paper where the assumption was made that the resistive force per molecule is inversely proportional to the number of molecules adsorbed. The equations developed here are not as satisfactory as those developed earlier and it is pointed out that this is a consequence of the fact that heats of adsorption have been found experimentally to decrease as the amount of adsorbed gas increases.

In an earlier paper (1) we have shown how the equation

$$\frac{\partial p}{\partial x} - A_x = 0, \quad (1)$$

where p is a pressure and A_x a resistive or frictional force, may be applied to the diffusion of adsorbed gases through solids and for such diffusion we concluded that the operative pressure is the spreading pressure, ϕ , of the adsorbed layer. In the second paper (2) the use of Equation (1) was extended to all types of diffusion and it was shown that, in general, the operative pressure can be determined from the thermodynamic relation

$$\frac{\partial F}{\partial V} = -p, \quad (2)$$

where F is the free energy and V the volume. By the use of (2) the pressure can be determined independently of any physical picture of the relation between the diffusing gas and the solid. On the other hand, to determine the resistive force A_x it is necessary to assume some mechanism of flow for the diffusing substance. If, therefore, Equation (1) is accepted as the fundamental dynamical equation for diffusion, the application to any specific system resolves itself into evaluating A_x correctly in terms of the known physical properties of the system.

In the treatment of the diffusion of an adsorbed gas through a solid as originally given (1), it was assumed that the resistive force could be put equal to $-Cu$ where C is the coefficient of resistance and u is the resultant velocity with which the adsorbed layer moves over the surface of the solid. By this assumption we treat C , which is the resistive force per unit volume when the velocity is unity, as a constant independent of the amount of gas adsorbed.

¹ Manuscript received September 18, 1950.

Contribution from the Division of Physics, National Research Laboratories, Ottawa, Canada. Issued as N.R.C. No. 2504.

It was originally pointed out that taking the resistive force as independent of the amount of gas adsorbed was an assumption that would require experimental confirmation. The experimental results quoted in that paper did indicate that such an assumption was justified and the resulting equations were put forward as a solution to the problem.

In expanding the scope of Equation (1) to cover diffusion in liquid and solid solutions, it has become apparent that, in putting $A_x = -Cu$, we were not making the assumption that was intended. Our intention had been to make the resistance per molecule, rather than per cc., independent of the amount of gas adsorbed but, owing to a wrong interpretation of the fundamental equation, we have not obtained that result. In this paper, therefore, we shall develop the equations as originally intended and compare them with experiment and with the earlier equations.

On the assumption that the resistive force per molecule is a constant independent of the number of molecules adsorbed we write $A_x = -C_m Nu$ where C_m is the coefficient of resistance per molecule and N is the total number of molecules adsorbed per cc. The fundamental equation becomes

$$\frac{\partial \phi}{\partial x} + C_m Nu = 0. \quad (3)$$

Hence the number of molecules diffusing across unit area in unit time is given by

$$Nu = -\frac{1}{C_m} \frac{\partial \phi}{\partial x}. \quad (4)$$

We shall now apply this equation to different models of adsorption.

Diffusion of a Mobile Monolayer

For a mobile monolayer we have $\phi = \frac{NkT}{A}$ where T is the temperature, k is Boltzmann's constant, and A the area of surface per cc. Substituting in (4) we get

$$Nu = -\frac{kT}{C_m A} \frac{\partial N}{\partial x}. \quad (5)$$

It follows from this that, for a mobile monolayer, the diffusion is directly proportional to the concentration gradient. Thus the diffusion obeys Fick's law. Since a mobile monolayer is the two dimensional analogue of a perfect gas it is not surprising to find that Fick's law is applicable.

Diffusion of an Ideal Localized Monolayer (Langmuir Adsorption)

For adsorption that can be represented by a Langmuir isotherm the spreading pressure is given by the relation

$$\phi = \frac{N_s kT}{A} \log \frac{1}{1 - \theta}, \quad (6)$$

where N_s is the number of sites per cc. and $\theta = N/N_s$. Substituting in (4) we get

$$Nu = - \frac{N_s kT}{C_m A} \frac{\partial}{\partial x} \log \frac{N_s}{N_s - N}, \quad (7)$$

$$= - \frac{N_s kT}{C_m A} \cdot \frac{1}{N_s - N} \cdot \frac{\partial N}{\partial x}. \quad (8)$$

Comparing with Fick's law in the form $P = D_c \frac{\partial c}{\partial x}$ where c is the concentration and P the rate of permeation, we see that

$$D_c \sim \frac{1}{N_s - N} \sim \frac{1}{1 - \theta}. \quad (9)$$

This is the relation between the diffusion coefficient of Fick's law and the θ of Langmuir's equation.

It is interesting to find the expression for the differential coefficient in terms of the gas pressure. The Langmuir equation gives

$$\theta = \frac{N}{N_s} = \frac{ap}{1 + ap},$$

where a is a constant. Hence

$$\frac{\partial N}{\partial x} = \frac{N_s}{(1 + ap)^2} \frac{\partial p}{\partial x}$$

and the expression for the diffusion becomes

$$Nu = - \frac{N_s kT}{C_m A} \cdot \frac{1}{1 + ap} \cdot \frac{\partial p}{\partial x}. \quad (10)$$

Comparing this with a diffusion equation in the form $P = - D_p \frac{\partial p}{\partial x}$, we see that

$$D_p \sim \frac{1}{1 + ap}. \quad (11)$$

It is apparent, therefore, that D_p becomes less as p increases and when ap is large compared to unity D_p finally approaches zero. Such behavior is to be expected from the fact that Langmuir adsorption approaches saturation as p is increased.

In practice it is always an integrated value of the diffusion that is measured. In order to compare theory with experiment, we assume that the diffusion takes place across a barrier of thickness L when there is a concentration $N = N_0$ at $x = 0$ and $N = N_L$ at $x = L$. To find the number of molecules diffusing from surface $x = 0$ to $x = L$ we integrate (7) remembering that in steady flow, by the equation of continuity, Nu is a constant. Then

$$NuL = \frac{N_s kT}{C_m A} \left[\log \frac{N_s}{N_s - N} \right]_{N=N_L}^{N=N_0} \quad (12)$$

and

$$Nu = \frac{N_s kT}{C_m AL} \log \frac{N_s - N_L}{N_s - N_0}.$$

In most experiments the pressure on one side of the barrier is zero so we take $N_L = 0$. Therefore

$$\begin{aligned} Nu &= \frac{N_s kT}{C_m AL} \log \frac{N_s}{N_s - N_0}, \\ &= \frac{N_s kT}{C_m AL} \log \frac{1}{1 - \theta_0}, \\ &= \frac{N_s kT}{C_m AL} \log (1 + ap_0). \end{aligned} \quad (13)$$

To test this equation we compare it with the measurements of Smithells and Ransley (7) for the diffusion of hydrogen through copper at 450°C. These measurements were also used as a test of the equation developed in the earlier

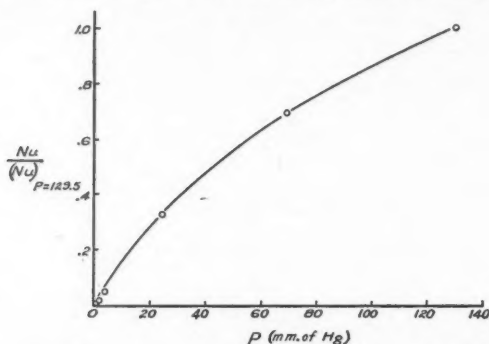


FIG. 1. Comparison of theoretical curve with the measurements of Smithells and Ransley for the diffusion of hydrogen through copper at 450°C. O Measurements.

paper. The agreement between theory and experiment is shown in Fig. 1. To calculate the curve, a has been taken as equal to 0.025 and the curve has been made to agree with the experimental results at the highest pressure measured, that is, at $p_0 = 129.5$. With these adjustments, the theoretical curves can be said to explain satisfactorily the experimental results but it should be noted that this equation does not give the reversed curvature suggested by the measurements at very low pressures and reproduced by the equation developed in the first paper (1). On the basis of these measurements it seems that the original equation is slightly more satisfactory.

As pointed out in the earlier paper the real test of these equations will come only when it is possible to measure a independently by adsorption measure-

ments; at present no adsorption isotherms are available at the temperatures at which diffusion measurements are possible. It is not surprising that both equations are satisfactory explanations of these experimental results since they differ only at low pressures where the measurements are hardly definite enough to discriminate between them.

Diffusion of a Gas Obeying the BET Isotherm

The spreading pressure ϕ of a gas adsorbed according to the equation of Brunauer, Emmett, and Teller (the BET equation) is given by

$$\phi = \frac{BkT}{A} \log \frac{B}{B-X}, \quad (14)$$

where B is the total number of localized sites and X is the number of these sites that are occupied. Substituting this value for ϕ in Equation (4), we get

$$\begin{aligned} Nu &= -\frac{BkT}{C_m A} \frac{\partial}{\partial x} \log \frac{B}{B-X}, \\ &= -\frac{kT}{C_m A} \frac{B}{B-X} \frac{\partial X}{\partial x}. \end{aligned} \quad (15)$$

To eliminate X we use the relation

$$(N-X)(B-X) = \beta X^2, \quad (16)$$

where N is the total number of molecules adsorbed and β is a constant. We get

$$Nu = -\frac{kT}{C_m A} \frac{B}{\sqrt{(N-B)^2 + 4\beta NB}} \cdot \frac{\partial N}{\partial x}, \quad (17)$$

which can be written

$$Nu = -\frac{kT}{C_m A} \cdot \frac{1}{\sqrt{\left(\frac{N}{B} - 1\right)^2 + 4\beta \frac{N}{B}}} \cdot \frac{\partial N}{\partial x}. \quad (18)$$

In practice the gas pressure p is invariably measured rather than the concentration and to express the diffusion in terms of the pressure we have the BET equation in the form

$$\frac{N}{B} = \frac{cP}{(1-P)(1-P+cP)}, \quad (19)$$

where $c = \frac{1}{\beta}$ and $P = \frac{p}{p_s}$; p_s is the saturation vapor pressure.

$$\text{From (18)} \quad \frac{\partial N}{\partial x} = \frac{1-P^2+cP^2}{[(1-P)(1-P+cP)]^2} \cdot \frac{\partial P}{\partial x} \quad (20)$$

so that the equation of diffusion in terms of the pressure becomes

$$Nu = -\frac{cBkT}{C_m A} \cdot \frac{1}{\sqrt{\left[\frac{cP}{(1-P)(1-P+cP)} - 1\right]^2 + \frac{4P}{(1-P)(1-P+cP)}}} \cdot \frac{1-P^2+cP^2}{[(1-P)(1-P+cP)]^2} \cdot \frac{\partial P}{\partial x} \quad (21)$$

Thus if we write the equation of diffusion in the form

$$Nu = -D_P \frac{\partial P}{\partial x}$$

we have

$$D_P \sim \frac{1}{\sqrt{\left[\frac{cP}{(1-P)(1-P+cP)} - 1\right]^2 + \frac{4P}{(1-P)(1-P+cP)}}} \cdot \frac{1-P^2+cP^2}{[(1-P)(1-P+cP)]^2} \quad (22)$$

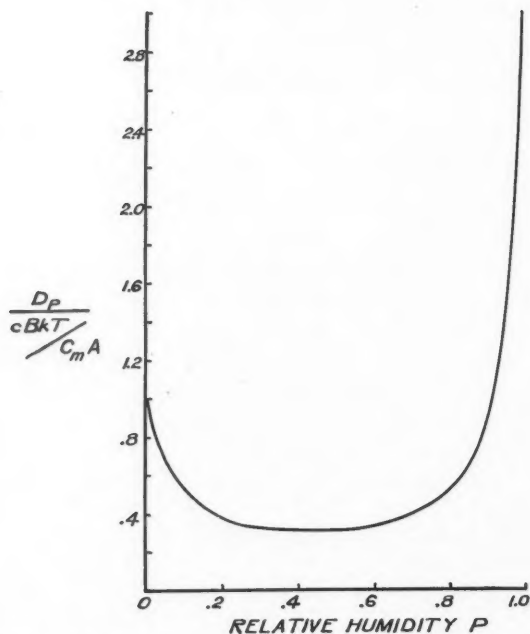


FIG. 2. Variation of differential coefficient D_P with pressure when the adsorption follows the BET isotherm.

In Fig. 2 we have plotted $D_p \left| \frac{cBkT}{C_m A} \right|$ against P . The curve shows that D_p decreases at low humidities from its value at zero pressure, passes through a minimum and then rises very sharply as the pressure approaches saturation. We see that at low pressures the diffusion is similar to that obtained for Langmuir adsorption and this is to be expected since the BET isotherm at low pressures follows the Langmuir isotherm. At high pressures, however, the diffusion increases to infinity following the BET isotherm which gives infinite adsorption at saturation pressures.

Again, to obtain a comparison with measurements, it is necessary to use an integrated form of the equation. Putting

$$(N - B)^2 + 4\beta NB \equiv Y \quad (23)$$

we can write Equation (17) in the form

$$Nu = - \frac{BkT}{C_m A} \cdot \frac{1}{\sqrt{Y}} \cdot \frac{\partial N}{\partial x}, \quad (24)$$

$$= - \frac{BkT}{C_m A} \frac{\partial}{\partial x} \log [\sqrt{Y} + N + (2\beta - 1)B]. \quad (25)$$

If we consider again the diffusion through a barrier of thickness L with $N = N_0$ at $x = 0$ and $N = N_L$ at $x = L$, we can integrate (25) remembering that in steady flow Nu is constant. Then

$$Nu = \frac{BkT}{C_m AL} \log [\sqrt{Y} + N + (2\beta - 1)B]_{N=N_L}^{N=N_0}. \quad (26)$$

To test this equation we make use of the measurements of King (5) on the diffusion of water vapor through keratin. King gives a curve showing the variation of rate of flow of water vapor through horn keratin with change of relative humidity on the high pressure side. In addition he gives the adsorption isotherm for the keratin so that it is possible to evaluate B and β directly from the BET equation.

In Fig. 3 the theoretical isotherm of the BET equation using $B = 5.66$ and $C = \frac{1}{\beta} = 17.67$ is compared with the experimental isotherm of King. Above a relative humidity of 0.5, the BET theory gives a higher adsorption than is found experimentally but below 0.5 the agreement is satisfactory.

The experimental results of King for the diffusion of water vapor through keratin are reproduced in Fig. 4 along with the diffusion calculated according to Equation (26) (full line curve). The theoretical curve has been made to agree with the measurements at a relative humidity of 0.5. Although below 0.5 relative humidity the same general trend can be seen in theory and observation, above this value the calculated results rise much too slowly. Although it is in

this region that the BET equation deviates from the experimental isotherm, it does not seem possible to attribute the difference between the theoretical curve and the measurements wholly to the inadequacy of the adsorption theory. It would appear that the BET theory would make the diffusion higher rather than lower at the high relative humidities. We can only conclude that Equation (26) is not a satisfactory explanation of the diffusion of water vapor through keratin.

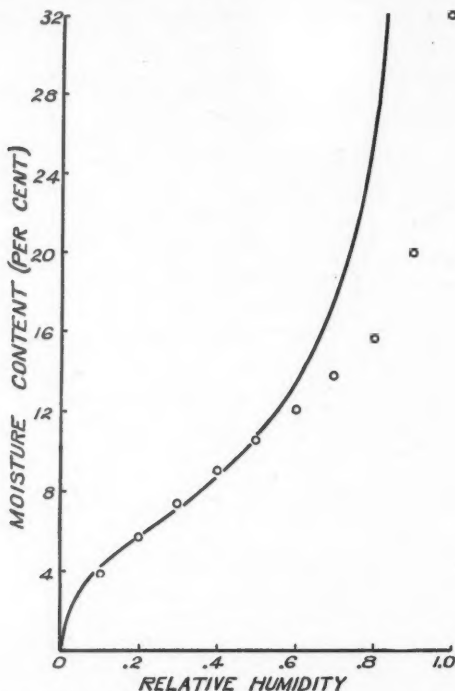


FIG. 3. Comparison of the BET isotherm with the measurements of King for the adsorption of water vapor by keratin. \circ Measurements. Curve calculated with $B = 5.66$ and $C = 17.67$.

Since the agreement between Equation (26) and the experiments of King cannot be considered satisfactory, it is interesting to revert to the equation developed in the earlier paper. That equation was based on the assumption that $A_x = -Cu$. The equivalent equation to (26) has the form

$$Nu = \frac{BkT}{CAL} \left\{ \sqrt{Y} - B(2\beta - 1) \log [\sqrt{Y} + N + B(2\beta - 1)] \right\} \Big|_{N=N_L}^{N=N_0} \quad (27)$$

When the data of King for the adsorption of water by keratin are applied to this equation we obtain the dotted curve of Fig. 4. This curve is in better agreement with the experimental results but still gives values for the diffusion

that are too low at high humidities. However, the fit with measurements at relative humidities below 0.5 is so good and the general shape of the curve follows so closely that of the measurements that it is reasonable to believe that this theory is very close to the true picture.

Since for both the Langmuir and the BET adsorption, the earlier theory, putting $A_x = -Cu$, is in better agreement with experiment than the theory developed in this paper, where $A_x = -C_mNu$, it is necessary to consider what

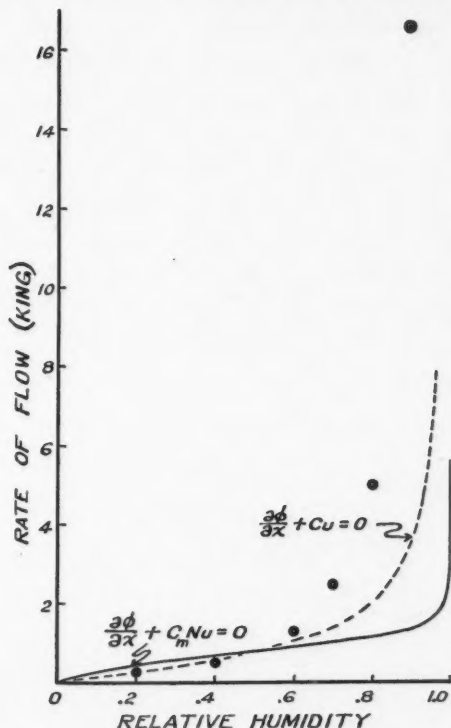


FIG. 4. Comparison of theory with the measurements of King for the rate of flow of water vapor through keratin. ● Measurements. Full curve calculated from equation $\frac{\partial \phi}{\partial x} + C_mNu = 0$; dotted curve calculated from equation $\frac{\partial \phi}{\partial x} + Cu = 0$.

this implies. Since A_x is the resistance per unit volume, the relation $A_x = -C_mNu$ is equivalent to the assumption that a molecule contributes a resistance to the flow that is independent of the number of molecules adsorbed whether it be large or small. This is the same thing as assuming that molecules adsorbed at low pressures are bound to the adsorbing surface with

the same energy as those molecules adsorbed near saturation. That this is not so is a fundamental assumption of the BET theory and it is not surprising, therefore, that the equation deduced from this assumption has failed to agree with experiment.

On the other hand, the assumption that $A_x = -Cu$ is equivalent to saying that the resistance per cc. is independent of the number of adsorbed molecules. The contribution made by each molecule on this picture is equal to C/N and is therefore inversely proportional to the number of molecules adsorbed. This amounts to the assumption that the binding force between adsorbed molecule and absorbing surface varies as $\frac{1}{N}$. There have been several estimates of the variation of heat of adsorption with number of molecules adsorbed (*cf.* for instance, Babbitt (3), Guthrie (4), Miller (6)) and all have found that the heat of adsorption decreases from a high value at zero adsorption to a value at saturation approaching that of the heat of liquefaction of the gas. In general, the curves relating heat of adsorption to number of molecules adsorbed are of such shape that an inverse relation between heat of adsorption and number of molecules adsorbed would be a fair approximation to these curves. This, then, is, without doubt, the reason why such good agreement is found for diffusion when the assumption is made that $A_x = -Cu$. On the basis of Fig. 4, however, it is apparent that in BET adsorption the molecules at high moisture contents are even more mobile than is required by the relation $C_m \sim 1/N$.

References

1. BABBITT, J. D. Can. J. Research, A, 28: 449. 1950.
2. BABBITT, J. D. Can. J. Phys. 29: 427. 1951.
3. BABBITT, J. D. Can. J. Research, A, 20: 143. 1942.
4. GUTHRIE, J. C. J. Textile Inst. 40: T, 489. 1949.
5. KING, G. Trans. Faraday Soc. 41: 479. 1945.
6. MILLER, A. R. The adsorption of gases on solids. Cambridge University Press, London. 1949.
7. SMITHELLS, C. J. and RANSLEY, C. E. Proc. Roy. Soc. (London), A, 150: 172. 1935.

AN ESTIMATE OF THE RELATIVE HELIUM CONTENT OF THE EARLY-TYPE STARS¹

BY ANNE B. UNDERHILL

Abstract

The relative abundance of hydrogen to helium in O-type atmospheres is estimated by a comparison of the computed strength of the HeII line $\lambda 4541$ with observed values. The line profile of $\lambda 4541$ in an O9.5 V and in an O5 V model atmosphere is computed when the ratio H:He = 5.67 and when it is 13.33, and the equivalent width of the line derived. Observations of the strength of $\lambda 4541$ in the spectra of the O9 V star, 10 Lacertae, and the O9.5 star, σ Orionis, are given, and the observed equivalent width of $\lambda 4541$ in a real stellar atmosphere of spectral type corresponding to the O9.5 V model atmosphere is estimated. A comparison of the observed strength of HeII $\lambda 4541$ with the computations indicates that the abundance ratio H:He lies in the range 20 to 25, that is, the relative abundance of helium is reduced by a factor of 1/3 to 1/2 from the amounts estimated by several authors by other spectroscopic methods. It is shown that the O9.5 V model atmosphere gives a fair representation of a late O-type main-sequence atmosphere, and that the computations can best be reconciled with observation by reducing the relative abundance of helium to about 4 or 5% by number, the rest of the stellar atmosphere being effectively pure hydrogen. The computations for the O5 model atmosphere are not in agreement with this result, but reasons are given for believing that the O5 model atmosphere does not give a good representation of an early O-type stellar atmosphere, and that consequently the results obtained from this model atmosphere are not of equal weight to those obtained from the O9.5 model atmosphere.

Introduction

It is a problem of general interest to derive the relative abundance of helium to hydrogen in stellar atmospheres. One way of doing this is by means of model atmospheres. Once a model atmosphere, calculated under certain well defined assumptions, has been obtained, the strength of any spectral line produced by it may be computed as a function of the abundance of the carrier atom in the atmosphere. A plot of the computed equivalent width of the line against abundance defines a theoretical curve of growth for the line. Comparison with the observed strength of the line in a real stellar atmosphere corresponding in spectral type to the model gives an estimate of the abundance of the element.

In stellar atmospheres the line $\lambda 4541$ of HeII will be broadened by Stark effect in the same way as are the hydrogen lines. A statistical theory of the Stark broadening of the hydrogen lines has been worked out by Verweij (16), and we shall use this theory. Following Unsöld (13), we shall consider only the Stark broadening caused by protons. Neglect of the Stark effect due to electrons means that the broadening by Stark effect of the line $\lambda 4541$ will be underestimated and that as a result the relative abundance of helium derived will, if anything, be a little too large.

We have reviewed the statistical theory of the Stark effect of hydrogen in a previous publication (10), and we find (10, Equation (15)) for a line in which

¹ Manuscript received May 22, 1951.

Contribution from the Dominion Astrophysical Observatory, Victoria, B.C.

the Stark broadening of each component is greater than the broadening due to the thermal motion of the atoms and to radiation damping, that the line-absorption coefficient per gram of star material corrected for stimulated emission, is

$$l_\nu d\nu = N^*(1 - e^{-h\nu/kT}) \left[\frac{\pi e^2 \lambda^2}{mc^2 F_0} f_\pm S(a) + \frac{\sqrt{\pi} e^2 f_0}{mc \Delta\nu_D} H(a, \nu) \right] d\nu. \quad (1)$$

Here N^* is the number of atoms per gram of star material excited to the lower level of the line, F_0 is the normal field strength in e.s.u., f_\pm is the part of the total f -value of the line due to the Stark components which are displaced in an electric field, f_0 is the part of the total f -value due to components of zero-displacement, $S(a)$ is the function introduced by Verweij (16) to describe the Stark broadening of a hydrogen line (a being the displacement expressed as $\Delta\lambda/F_0$), the function $H(a, \nu)$ gives the shape of the line under the action of thermal broadening and damping of a dispersion type, and $\Delta\nu_D$ is the half-width of the Doppler core of the line. To use Equation (1) for $\lambda 4541$ of HeII, $S(a)$, f_\pm , and f_0 must be obtained. The function $H(a, \nu)$ may be found from the tabulation by Harris (6).

The definition of the function $S(a)$ is

$$S(a) = \frac{1}{n'^2 f_\pm} \sum_k \frac{f_k}{c_k} W\left(\frac{a}{c_k}\right), \quad (2)$$

where the summation is carried out over all the Stark components of positive and negative displacement. Here f_k is the f -value of the Stark component k , and is given by

$$f_k = \frac{8\pi^2 m \nu}{3h} |z_k|^2, \quad (3)$$

where $|z_k|^2$ is the square of the matrix element for the transition in question, and $W(a/c_k)$ is the Holtzmark distribution function, which is tabulated, for instance, by Verweij (16). The quantity c_k is a numerical factor depending on the quantum numbers of the Stark component,

$$c_k = \frac{3h\lambda^2}{8\pi^2 m e c^2 Z} X_k,$$

where Z is the nuclear charge, 2 in this case, and

$$X_k = n(n_2 - n_1) - n'(n_2' - n_1'). \quad (4)$$

In Equation (4), n , is the chief quantum number of the upper level of the line and n_1 and n_2 are the parabolic quantum numbers of a particular upper sub-level; n' , n_1' , n_2' are similar quantities for the lower level of the line. For $\lambda 4541$ of HeII $n' = 4$, $n = 9$. The quantity f_\pm is defined by

$$f_\pm = \frac{1}{n'^2} \sum_k f_k, \quad (5)$$

where the summation is extended over *all* Stark components of positive and negative displacement. The quantity f_0 is defined in a similar manner, the sum being extended only over components of zero displacement.

We have computed the necessary matrix elements and evaluated the f_k according to Equation (3), using the formulae for the matrix elements given by Gordon (5) and incorporating the correction of the misprint previously pointed out (10). From this material the function $S(\alpha)$ for $\lambda 4541$ of HeII has been computed according to Equation (2). This function, which is given in

TABLE I
THE FUNCTION $S(\alpha)$ FOR $\lambda 4541$ OF HeII

α Å./e.s.u.	$S(\alpha)$	α Å./e.s.u.	$S(\alpha)$
0.000	0.000×10^7	0.14	1.332×10^8
0.005	8.105×10^7	0.18	1.034×10^8
0.01	1.322×10^8	0.22	7.854×10^7
0.02	1.856×10^8	0.26	5.767×10^7
0.03	1.604×10^8	0.30	4.174×10^7
0.04	1.593×10^8	0.40	2.420×10^7
0.06	1.692×10^8	0.60	5.936×10^6
0.08	1.754×10^8	0.80	2.706×10^6
0.10	1.648×10^8	1.00	1.477×10^6
0.12	1.502×10^8	1.50	5.00×10^5
0.14	1.332×10^8	2.00	2.11×10^5

Table I, has generally the same shape as the similar functions computed by Verweij (16) for $H\alpha$, $H\beta$, $H\gamma$, and $H\delta$. We find that $f_{\pm} = 0.018109$ and $f_0 = 0.000539$.

In the wave-length range for which the collisional theory of line broadening is valid, that is, within the distance $\Delta\lambda_0$ of the line center, the line absorption coefficient is

$$l_\nu d\nu = N^*(1 - e^{-h\nu/kT}) \frac{\sqrt{\pi} e^2}{mc} \frac{f}{\Delta\nu_D} H(a, \nu) d\nu, \quad (6)$$

where f is the total f -value of the line, and the value of the damping constant γ (which enters into the quantity a), includes the effect of collisions. For the late O-type stars, the distance $\Delta\lambda_0$ is, following the arguments of Unsöld (14), found to be almost 0.2 Å . Hence we may use the statistical theory of line broadening over effectively the whole line profile.

The Model Atmosphere Computations

The method of computing line profiles has been described elsewhere (11, 12). Using the form of the line-absorption coefficient given above, Equations (1) and (6), we can obtain l_ν at each level in the model atmosphere where T and P_e are known. The ionization and excitation of the helium ions is assumed to take place in accordance with a state of local thermodynamic equilibrium.

There are two O-type model atmospheres for which we shall compute the strength of the line $\lambda 4541$. The first (11) has $T_0 = 25,200^\circ\text{K.}$ and $\log g = 4.200$. The integrated flux from this model corresponds to an effective temperature of $36,800^\circ$ and the computed relative strengths of lines of CII/CIII and SiIII/SiIV place the spectral type of this model as O9.5 V. The second model atmosphere (12) has $T_0 = 35,000^\circ$ and $\log g = 4.200$. The integrated flux from this model gives an effective temperature of $44,600^\circ$ and the spectral type, according to the absolute strength of the line $H\gamma$, is about O5. The structure of both models has been computed using the equation of mechanical equilibrium (allowance being made for the radiation pressure gradient), and in both models the stellar atmosphere is assumed to be composed of hydrogen and helium only, in the ratio H: He = 5.67 by number.

The structure of the model atmosphere, that is the temperature, electron pressure, and total pressure at any optical depth τ , depends upon the composition of the model atmosphere; hence, strictly, we should compute a new model atmosphere for each value of the ratio H: He assumed. However, since preliminary calculations showed that the structure was not very sensitive to the abundance of helium assumed, we have assumed that the structure of the model atmospheres (11, 12) found for the case H: He = 5.67 gives an adequate representation of the structure when the ratio H: He varies somewhat from 5.67.

In Table II the computed profile of $\lambda 4541$ in the O9.5 V model atmosphere is given for the case H: He = 5.67 and for H: He = 13.33. The equivalent

TABLE II
THE LINE PROFILE OF $\lambda 4541$ IN THE MODEL ATMOSPHERES

R _v in O9.5 model			R _v in O5 model		
$\Delta\lambda, \text{\AA.}$	H: He = 5.67	H: He = 13.33	$\Delta\lambda, \text{\AA.}$	H: He = 5.67	H: He = 13.33
0.0	0.778	0.808	0.0	0.848	0.844
0.4	0.872	0.906	0.5	0.863	0.873
0.7	0.891	0.917	1.0	0.895	0.902
1.0	0.906	0.928	2.0	0.942	0.960
2.0	0.940	0.951	5.0	0.991	0.994
5.0	0.977	0.981	10.0	0.997	0.997
10.0	0.992	0.992			

width in the first case is 0.81 equiv. \AA. , and in the second case is 0.65 equiv. \AA. Also given is the computed profile of $\lambda 4541$ for the O5 model atmosphere for these two cases. For the O5 model atmosphere the equivalent width is 0.71 equiv. \AA. when H: He = 5.67 and 0.58 equiv. \AA. when H: He = 13.33. In order to compute the line-absorption coefficient at $\Delta\lambda = 0.0 \text{\AA.}$ and to compute the part of the line-absorption coefficient due to the Stark components of zero displacement, it was assumed, in accord with previous results, that the damping constant, γ , was 10 times the classical damping constant. Since the exact value used for the damping constant is immaterial, it was not considered worthwhile

to evaluate this constant in detail; instead a value of the expected correct order of magnitude was taken, that is $\gamma = 5.4 \times 10^8$.

The Relative Abundance of Helium to Hydrogen

The computed equivalent widths for the O9.5 V model atmosphere are plotted in Fig. 1 to give a curve of growth. The ratio H: He = 5.67, found by Unsöld (15) for the B0 V star τ Scorpii and also the value H: He = 10, found by Aller (1) for 10 Lacertae and by Bowen and Wyse (4) and Aller and Menzel (2) from a study of the relative intensities of the hydrogen and helium emission lines in planetary nebulae, are indicated. The curve of growth has been drawn as a straight line through the two computed points.

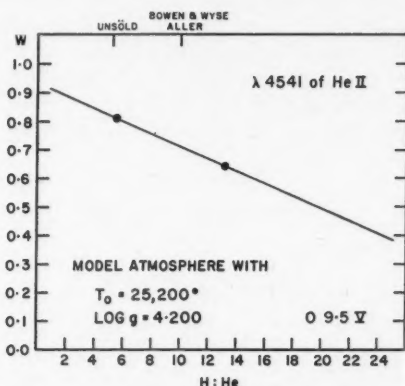


FIG. 1. The theoretical curve of growth of $\lambda 4541$ of He II in an O9.5 V model atmosphere. The units of equivalent width, W , are equivalent angstroms. The relative abundance H: He is by number.

We shall now obtain an estimate of the observed equivalent width of $\lambda 4541$ in O9.5 V spectra and, by interpolation in Fig. 1, derive the abundance of helium in late O-type stellar atmospheres.

The equivalent width of $\lambda 4541$ of He II has been measured on five plates of 10 Lacertae, O9 V, the measurements being made by integrating the area of the line on intensity tracings of the spectral region surrounding $\lambda 4541$. The results of these measurements and of measurements on four plates of σ Orionis, O9.5 V, are given in Table III. The prism spectrograph used and the dispersion at $\lambda 4541$ on each plate are indicated. Each measure is weighted according to the dispersion and the quality of the spectrogram. The weighted mean equivalent width is given at the bottom of Table III for each star.

For 10 Lacertae we have found the following published measures of the equivalent width of $\lambda 4541$. Weights have been assigned to these measures

Author	$W(\lambda 4541)$	Weight
L. H. Aller (1)	0.37 equiv. Å.	1
E. G. Williams (17)	0.46	1
R. M. Petrie (9)	0.68	2
E. A. Baker (3)	0.79	2

Weighted mean $W(\lambda 4541)$	0.63	6
---------------------------------	------	---

according to their seeming reliability and a weighted mean equivalent width has been obtained. This value, 0.63 equiv. Å. is rather smaller than the value derived from the new measures in Table III, and probably this difference may be accounted for, at least in part, by the neglect of the extended, shallow Stark wings of $\lambda 4541$ (see below) by the above authors, although the method adopted by Baker should display these wings. The discrepancy is upsetting, but a study of our high-dispersion line profiles, and of the computed shape of the line $\lambda 4541$, encourages us to give considerable weight to our measurements. The weighted mean equivalent width of $\lambda 4541$ in 10 Lacertae from all the measures is 0.85 equiv. Å. The line $\lambda 4541$ is not present (15) in the B0 V star, τ Scorpii, hence it is likely that the equivalent width of $\lambda 4541$ in a star

TABLE III
THE MEASURED EQUIVALENT WIDTH OF $\lambda 4541$

10 Lacertae, O9V					σ Orionis, O9.5V				
Plate No.	Spectro-graph	Dispersion, Å./mm.	Weight	W , Equiv. Å.	Plate No.	Spectro-graph	Dispersion, Å./mm.	Weight	W , Equiv. Å.
41308	III	13.4	1	0.65	41680	IM	36.0	1	0.34
41310	III	13.4	3	1.21	41681	IM	36.0	1	0.31
38919	IL	26.4	2	0.88	41682	IM	36.0	1	0.24
41577	IM	36.0	1	1.00	41683	IM	36.0	1	0.23
41595	IM	36.0	1	1.14					
Mean equivalent width			8	1.02	Mean equivalent width			4	0.28

of spectral type halfway between that of 10 Lacertae and τ Scorpii will be about 0.5 equiv. Å. The observed equivalent width in σ Orionis, O9.5 V, from Table III, is 0.28 equiv. Å. Thus it seems likely that the observed equivalent width of $\lambda 4541$ in a real stellar atmosphere corresponding to our O9.5 V model atmosphere is about 0.4 equiv. Å. It is fairly certain that the equivalent width will not be much larger than 0.5 equiv. Å. or much smaller than about 0.3 equiv. Å. When we enter Fig. 1 with these values, we see that the ratio H: He must be greater than 20 and probably lies between about 20 and 25. That is, the present result indicates helium to be only 1/3 to 1/2 as abundant as previous estimates by Unsöld (15), Aller (1), Bowen and Wyse (4), and Aller and Menzel (2) have indicated, and as the reports in the Transactions of the International Astronomical Union (8) have suggested.

The observed profile of $\lambda 4541$ in 10 Lacertae, O9 V, and in σ Orionis, O9.5 V, and the computed profile in the O9.5 V model atmosphere, H: He = 13.3, are shown in Fig. 2. The intensity tracing for 10 Lacertae is from a two-prism plate magnified 80 \times , while that for σ Orionis is from a one-prism plate, magnified about 220 \times , to make the dispersion in the diagram the same for both tracings. A mean profile from two or more tracings is drawn in each case, although the grain is from one plate. The shallow, extended Stark wings of the computed profile can be clearly traced on the profile of $\lambda 4541$ in 10 Lacertae where the line is strong, and their existence is suggested in the profile for σ Orionis where the line is weak and where the grain size and low dispersion make it difficult to obtain a meaningful line profile. It is obvious that the computed line is too strong for spectral type O9.5 V. Thus, either we have

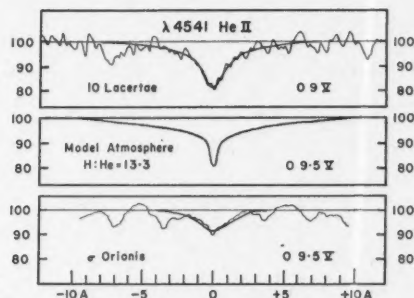


FIG. 2. The observed profile of HeII $\lambda 4541$ in 10 Lacertae, O9 V, and in σ Orionis, O9.5 V, and the computed line profile when H: He = 13.3.

assumed too much helium in the model atmosphere, the conclusion we reached above, or a wrong spectral type has been assigned to the model atmosphere, or for some reason inherent in the assumptions and calculations, the model atmosphere does not closely represent a real atmosphere.

In order to check the spectral classification of the model atmosphere, the equivalent widths of the CII, CIII, and SiIV lines used to classify the model atmosphere (11) have been measured. Values are found in essential agreement with those of Aller (1), which were used in the previous spectral classification (11) of the model atmosphere. There is little reason to believe that the model atmosphere is as early as 10 Lacertae in spectral type. Rather, it is halfway between 10 Lacertae and τ Scorpii, that is, about O9.5 V. Thus, since the spectral classification of the model atmosphere appears to be correct, it is likely that we have assumed too large an abundance of helium, and that the true H: He ratio is in the vicinity of 20 or 25.

There remains the question, only mentioned above, of whether the structure of a model with a small amount of helium only will be closely the same as that for a model with H: He = 5.67, and whether, indeed, the model represents a real

stellar atmosphere at all. In order to estimate the change in structure caused by a different admixture of helium, we have integrated, but not tested, for radiative equilibrium, a pure hydrogen model with the same temperature and surface gravity as the O9.5 V model atmosphere. An abbreviated summary of the results is given in Table IV. Both models have a boundary temperature $T_0 = 25,200^\circ$ and $\log g = 4.200$. It is assumed that the gas pressure is given by $P_g = 2P_e$, and that the temperature law is

$$T^4 = T_0^4 \left(1 + \frac{3}{2} \tau \right).$$

It is seen that the structures of the two models do not differ greatly, the electron pressure being reduced a little in the pure hydrogen model. The lowering of the electron pressure with the lessening of the helium content will tend to advance the second ionization of helium beyond the values used; however, the electric fields arising from protons that cause the Stark effect of the HeII lines increase slightly in a model with lessened helium content because of the increased abundance of hydrogen, hence there should be little change

TABLE IV

THE DEPENDENCE OF $\log P_e$ UPON HELIUM CONTENT IN THE O9.5 MODEL ATMOSPHERE*

τ	H: He = ∞ $\log P_e$	H: He = 5.67 $\log P_e$	$\frac{(P_e)H: He = 5.67}{(P_e)H: He = \infty}$	τ	H: He = ∞ $\log P_e$	H: He = 5.67 $\log P_e$	$\frac{(P_e)H: He = 5.67}{(P_e)H: He = \infty}$
0.00	1.820	1.820	1.00	1.00	3.152	3.189	1.09
0.05	2.364	2.364	1.00	1.40	3.276	3.318	1.10
0.10	2.513	2.513	1.00	2.20	3.474	3.517	1.10
0.20	2.617	2.625	1.02	2.60	3.545	3.591	1.11
0.50	2.911	2.936	1.05	3.00	3.606	3.657	1.12

* For these model atmospheres, $T_0 = 25,200^\circ$, $\log g = 4.200$, $T^4 = T_0^4 \left(1 + \frac{3}{2} \tau \right)$, and $P_g = 2P_e$.

in the value of the line-absorption coefficient of lines of HeII. We conclude that the approximation of using the structure of a model atmosphere computed with the ratio H: He = 5.67 to compute the strength of $\lambda 4541$ of HeII, when the abundance of helium is rather less than this, is valid and will not introduce serious errors into the theoretical curve of growth. We have confidence that the model atmosphere with $T_0 = 25,200^\circ$ and $\log g = 4.200$ gives a good representation of a real O9.5 V stellar atmosphere, since the strengths of the CII, CIII, SiIII, and SiIV lines computed for spectral classification (11) are all of the expected order of magnitude. Since these lines have a variety of excitation potentials, and since the ions concerned have a wide variety of ionization potentials, it seems unlikely that all the lines would have computed equivalent widths close to the observed equivalent widths if the structure of the model atmosphere did not give a fair representation of a real stellar atmosphere. For this model atmosphere, the effect of the radiation pressure on the mechanical

equilibrium is small; moreover the occurrence of sharp-lined late O-type stars such as 10 Lacertae suggests that main-sequence late O-type stars do have atmospheres in mechanical equilibrium as has been assumed. Thus, since the computations of the strength of the HeII line $\lambda 4541$ can be brought into accordance with observation by postulating a reduced abundance of helium, and since there are no indications that the model is not fairly representative of a real stellar atmosphere, we conclude that the model atmosphere computations presented here are evidence in favor of the ratio H: He being between 20 and about 25. That is, helium is present only to an amount of 4 to 5% by number, the rest of the stellar atmosphere being effectively pure hydrogen.

The recognition that the lines of the Pickering series of HeII have shallow, greatly extended Stark wings, Fig. 2, has some observational consequences. When a visual inspection of spectrograms is made, it will seem that in O stars with poor lines, the HeII lines, $\lambda 4541$ and $\lambda 4200$ in particular, will tend to disappear at earlier spectral types than they will in late O-type stars with good lines, for the eye is more influenced by the core of a line, than by wide-spread, shallow wings. Thus since a visual estimate of the relative strength of the HeII/HeI lines is used as a criterion of spectral classification, there may be a systematic error in the classification of O-type stars with poor lines to the effect that they are classified somewhat later than they should be. Also the shallow widespread wings of the Pickering lines explain why these lines are difficult to measure for radial velocity.

For the O5 model atmosphere (12) we have found that the equivalent width of $\lambda 4541$ is 0.71 equiv. Å. when H: He = 5.67 and 0.58 equiv. Å. when H: He = 13.33. Both these values are significantly less than the corresponding values for the O9.5 model atmosphere, a result that is in contradiction to the observed relative strengths of the HeII lines in O-type stellar atmospheres. Observationally it has been established (see for instance Ref. (9)) that the HeII lines do not reach maximum intensity until at least O5, and that the line $\lambda 4541$ increases in strength quite considerably from type O9.5 to type O5. Either the abundance of helium must be greater in early O-type stars than in late O-type stars, or our O5 model atmosphere does not give a good representation of a pure absorption O5 atmosphere.

According to Petrie (9) the equivalent width of $\lambda 4541$ should be expected to be at least 1.0 equiv. Å. at O5. If a curve of growth is constructed with the equivalent widths of $\lambda 4541$ computed above, as in Fig. 1, and this diagram entered at the point H: He = 20, which was found above to be a fair estimate of the hydrogen-helium ratio, it is found that $W(\lambda 4541)$ will be 0.4 equiv. Å. in the O5 model atmosphere, a value that is much too small. The relative strength of H γ to $\lambda 4541$ in the model atmosphere places the model as about O5 according to Petrie's (9) classification ratios. Such a spectral type for the model is consistent with the temperatures estimated for the early O-type stars from observation. The question then is, why are the calculated ionized helium lines so weak for the O5 model atmosphere?

It will be recalled that this model atmosphere (12) was assumed to be in strict mechanical equilibrium and to have $H:He = 5.67$ by number. To estimate the effect on the structure of the model of altering the helium abundance, we have integrated a pure hydrogen model with the same temperature law and surface gravity. A comparison of the resulting electron pressure in the two cases, $H:He = 5.67$ and $H:He = \infty$ is given in Table V. It is seen that the structure of this model is quite sensitive to the assumed abundance of helium and that the electron pressure is sensibly lowered as the abundance of helium is decreased. However, it is doubtful whether the changes due to differing helium abundance are large enough to account for the discrepancy found. It seems more likely that the absorption-line early O stars have atmos-

TABLE V
THE DEPENDENCE OF $\log P_e$ UPON HELIUM CONTENT IN THE O5 MODEL ATMOSPHERE*

τ	$H:He = \infty$ $\log P_e$	$H:He = 5.67$ $\log P_e$	$\frac{(P_e)H:He = 5.67}{(P_e)H:He = \infty}$	τ	$H:He = \infty$ $\log P_e$	$H:He = 5.67$ $\log P_e$	$\frac{(P_e)H:He = 5.67}{(P_e)H:He = \infty}$
0.00	1.831	1.954	1.33	1.00	3.505	3.585	1.20
0.05	2.676	2.769	1.24	1.40	3.623	3.702	1.20
0.10	2.854	2.939	1.22	2.20	3.826	3.890	1.16
0.20	2.966	3.049	1.21	2.60	3.907	3.968	1.15
0.50	3.272	3.350	1.20	3.00	3.980	4.038	1.14

* For these model atmospheres, $T_0 = 35,000^\circ$, $\log g = 4.200$, $T^4 = T_0^4 \left(1 + \frac{3}{2} \tau\right)$, and P_g is found as the appropriate function of P_e .

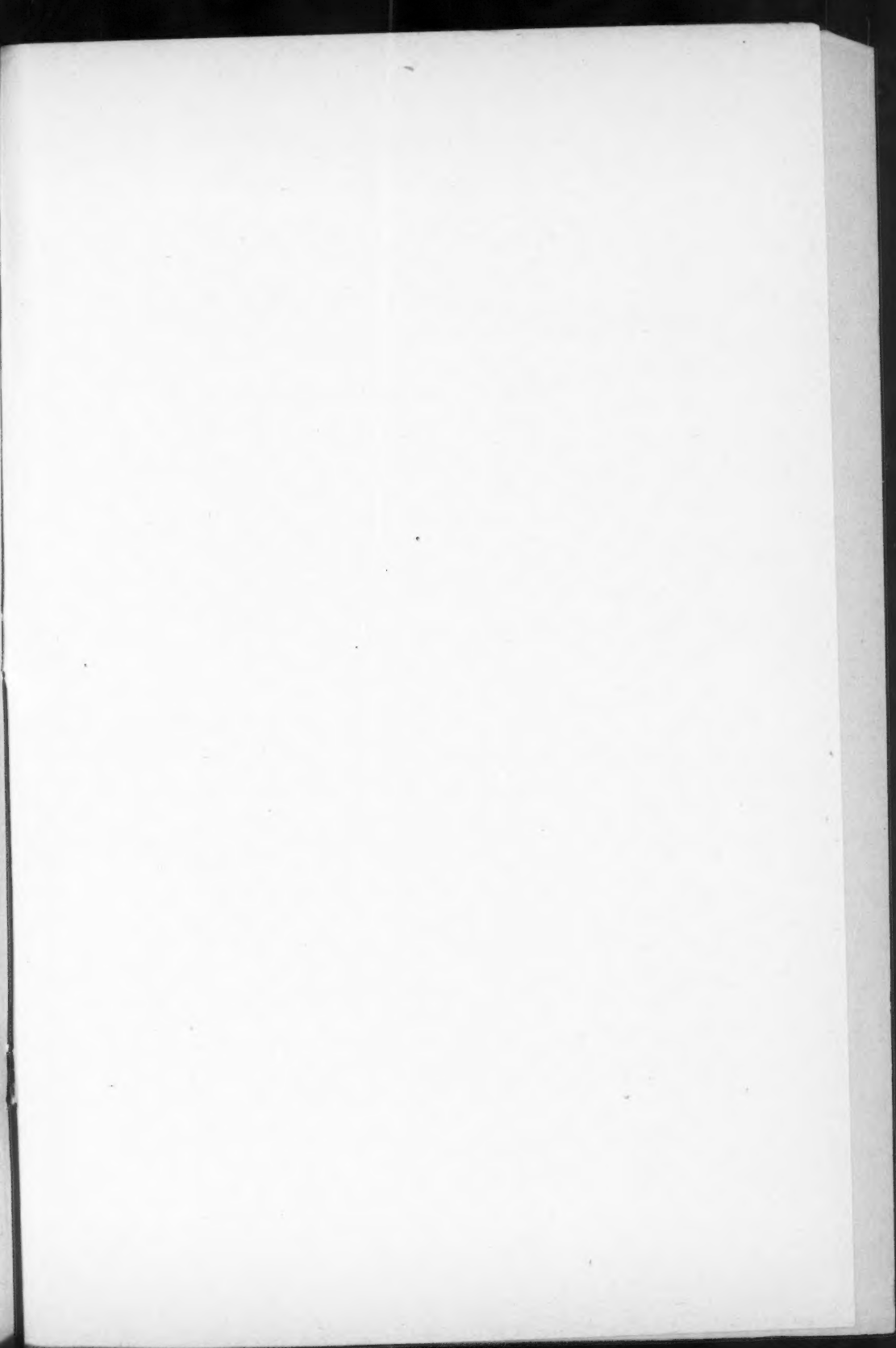
pheres in a state of turbulence. Then the equivalent widths of the absorption lines may be larger (7, 18) than those predicted from a stationary atmosphere in strict mechanical equilibrium. When the computations of the O5 model atmosphere (12) are examined, it is seen that the radiation pressure exceeds the gas pressure through most of the atmosphere. It seems likely that under these circumstances strict mechanical equilibrium, such as was forced on the model, may not exist, and that the O5 model atmosphere does not give a good representation of an early O-type stellar atmosphere. Thus the computed strength of the line $\lambda 4541$ of HeII in the O5 model atmosphere cannot safely be used to obtain an estimate of the relative abundance of helium to hydrogen in early type stars. Our conclusions should be based on the results for the O9.5 model atmosphere only. From the study of this model atmosphere we have concluded that the relative abundance by number of hydrogen to helium lies in the range 20 to 25.

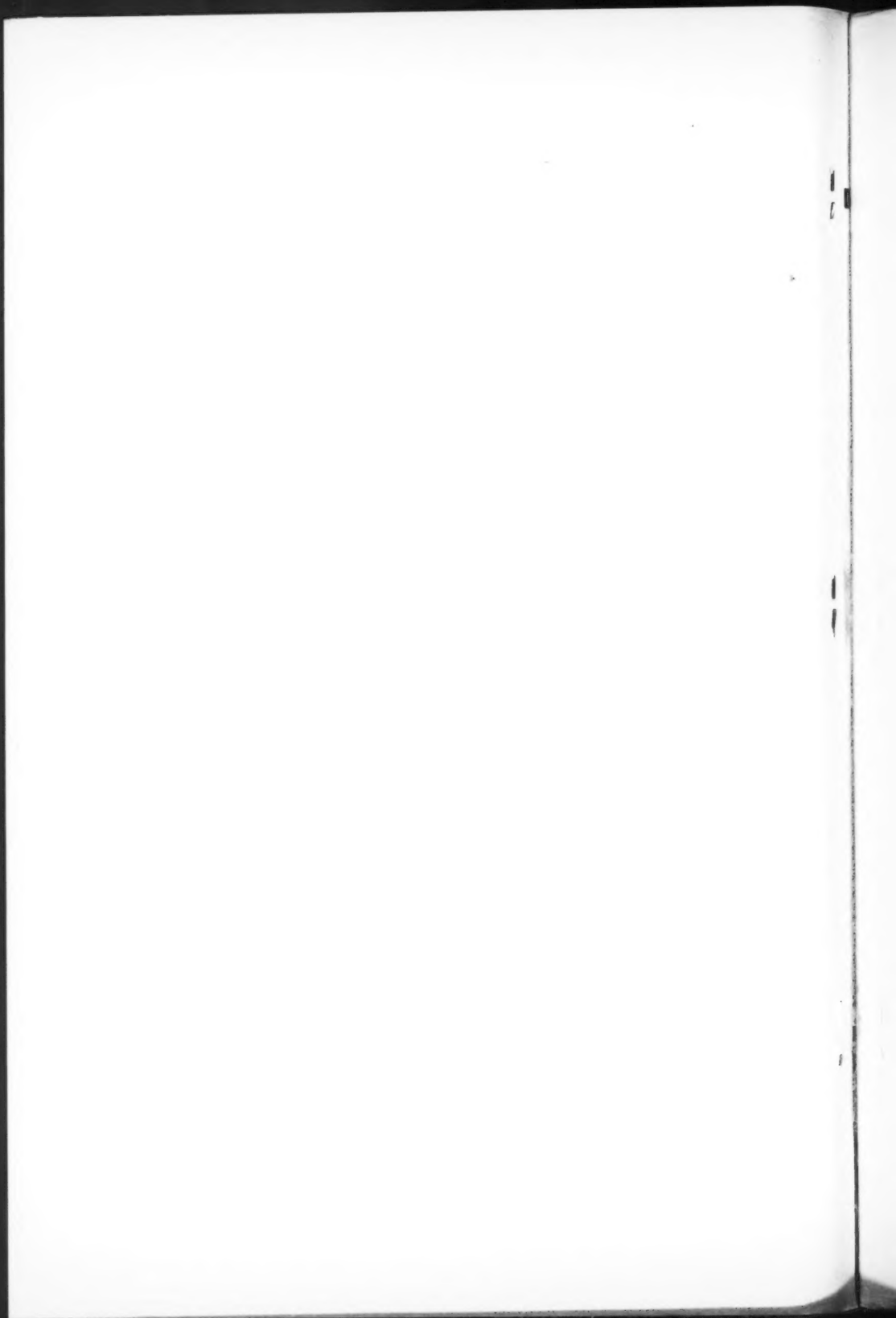
Acknowledgments

My sincere thanks are due to P. E. Argyle for his assistance in computing the matrix elements of the many Stark components of $\lambda 4541$, and to R. M. Petrie, A. McKellar, and J. G. Odgers for valuable discussions of the advantages and disadvantages of using model atmospheres to determine relative abundances.

References

1. ALLER, L. H. *Astrophys. J.* 104: 347. 1946.
2. ALLER, L. H. and MENZEL, D. H. *Astrophys. J.* 102: 239. 1945.
3. BAKER, E. A. *Pubs. Royal Observ. Edinburgh*, 1: 15. 1949.
4. BOWEN, I. S. and WYSE, A. B. *Lick Observ. Bull.* 19: 1. 1939.
5. GORDON, W. *Ann. Physik*, 2: 1031. 1929.
6. HARRIS, D. L. *Astrophys. J.* 108: 112. 1948.
7. HUANG, S. S. *Astrophys. J.* 112: 399, 418. 1950.
8. MINNAERT, M., UNSÖLD, A., and MENZEL, D. L. *Trans. Intern. Astron. U.* 7: 457. 1950.
9. PETRIE, R. M. *Pubs. Dominion Astrophys. Observ., Victoria, B.C.* 7: 321. 1947.
10. UNDERHILL, ANNE B. *Pubs. Dominion Astrophys. Observ., Victoria, B.C.* 8: 385. 1951.
11. UNDERHILL, ANNE B. *Pubs. Copenhagen Observ.* No. 151. 1950.
12. UNDERHILL, ANNE B. *Pubs. Dominion Astrophys. Observ., Victoria, B.C.* 8: 357. 1951.
13. UNSÖLD, A. *Vierteljahrsschr. Astron. Ges.* 78: 213. 1943. See especially p. 243 *et seq.*
14. UNSÖLD, A. *Vierteljahrsschr. Astron. Ges.* 78: 235. 1943.
15. UNSÖLD, A. *Z. Astrophys.* 21: 22. 1942.
16. VERWEIJ, S. *Pubs. Astron. Inst. Amsterdam.* No. 5. 1936.
17. WILLIAMS, E. G. *Astrophys. J.* 83: 279. 1936.
18. WRUBEL, M. H. *Astrophys. J.* 112: 424. 1950.





CANADIAN JOURNAL OF PHYSICS

Notice to Contributors

GENERAL: Manuscripts should be typewritten, double spaced, and the **original and one extra copy** submitted. Style, arrangement, spelling, and abbreviations should conform to the usage of this Journal. Names of all simple compounds, rather than their formulas, should be used in the text. Greek letters or unusual signs should be written plainly or explained by marginal notes. Superscripts and subscripts must be legible and carefully placed. Manuscripts should be carefully checked before being submitted, to reduce the need for changes after the type has been set. If authors require changes to be made after the type is set, they will be charged for changes that are considered to be excessive. **All pages, whether text, figures, or tables, should be numbered.**

ABSTRACT: An abstract of not more than about 200 words, indicating the scope of the work and the principal findings, is required.

ILLUSTRATIONS:

(i) **Line Drawings:** All lines should be of sufficient thickness to reproduce well. Drawings should be carefully made with India ink on white drawing paper, blue tracing linen, or co-ordinate paper **ruled in blue only**; any co-ordinate lines that are to appear in the reproduction should be ruled in black ink. Paper ruled in **green, yellow, or red should not be used** unless it is desired to have all the co-ordinate lines show. Lettering and numerals should be neatly done in India ink preferably with a stencil (**do not use typewriting**) and be of such size that they will be legible and not less than one millimeter in height when reproduced in a cut three inches wide. All experimental points should be carefully drawn with instruments. Illustrations need not be more than two or three times the size of the desired reproduction, but the ratio of height to width should conform with that of the type page. **The original drawings and one set of small but clear photographic copies are to be submitted.**

(ii) **Photographs:** Prints should be made on glossy paper, with strong contrasts; they should be trimmed to remove all extraneous material so that essential features only are shown. Photographs should be submitted **in duplicate**; if they are to be reproduced in groups, one set should be so arranged and mounted on cardboard with rubber cement; the duplicate set should be unmounted.

(iii) **General:** **The author's name, title of paper, and figure number should be written in the lower left hand corner (outside the illustration proper) of the sheets on which the illustrations appear.** Captions should not be written on the illustrations, but typed on a separate page of the manuscript. All figures (including each figure of the plates) should be numbered consecutively from 1 up (arabic numerals). **Each figure should be referred to in the text.** If authors desire to alter a cut, they will be charged for the new cut.

TABLES: Each table should be typed on a separate sheet. Titles should be given for all tables, which should be numbered in Roman numerals. Column heads should be brief and textual matter in tables confined to a minimum. **Each table should be referred to in the text.**

REFERENCES: These should be listed **alphabetically by authors' names, numbered in that order, and placed at the end of the paper.** The form of literature citation should be that used in this Journal. **Titles of papers should not be given.** The first page only of the references cited should be given. **All citations should be checked with the original articles.** Each citation should be referred to in the text by means of the key number.

REPRINTS: A total of 50 reprints of each paper, without covers, are supplied free to the authors. Additional reprints will be supplied according to a prescribed schedule of charges. On request, covers can be supplied at cost.

Approximate charges for reprints may be calculated from the number of printed pages, obtained by multiplying by 0.6 the number of manuscript pages (double-spaced typewritten sheets, 8½ in. by 11 in.) and making allowance for space occupied by line drawings and half-tones (not inserts). The cost per page is tabulated at the back of the reprint request form sent with the galley.

Manuscripts should be addressed: *Canadian Journal of Physics,*
National Research Council,
Ottawa, Canada.

Contents

	Page
Spectroscopic Study of Small Samples in a Hollow Cathode Discharge— <i>K. B. Newbound and F. H. Fish</i> - - - - -	357
The Effect of Ion Bombardment on the Emission from Oxide Coated Cathodes— <i>P. A. Redhead</i> - - - - -	362
The Macromosaic Structure of Tin Single Crystals— <i>E. Teghtsoonian and Bruce Chalmers</i>	370
Dendritic Growth in Lead— <i>F. Weinberg and Bruce Chalmers</i> - -	382
Theory of Multipole Radiations— <i>P. R. Wallace</i> - - - - -	393
Variation of Meteor Echo Rates with Radar System Parameters— <i>D. W. R. McKinley</i> - - - - -	403
A Unified Picture of Diffusion— <i>J. D. Babbitt</i> - - - - -	427
On the Diffusion of Adsorbed Gases Through Solids— <i>J. D. Babbitt</i>	437
An Estimate of the Relative Helium Content of the Early-type Stars— <i>Anne B. Underhill</i> - - - - -	447

